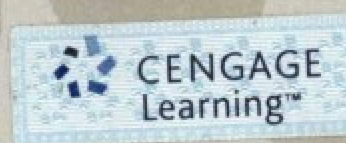
S H E H U I K E X U E Y A N J I U D E S I W E I Y A O S U

SHEHUI KEXUE
YANJIU
DE
SIWEI YAOSU

万卷方法

社会科学研究的思维要素

■肯尼斯·赫文 托德·多纳 著
■李涤非 潘 磊 译





重庆大学出版社
http://www.cqup.com.cn

S H E H U I K E X U E Y A N J I U D E S I W E I Y A O S U

SHEHUI KEXUE
YANJIU
DE
SIWEI YAOSU

有很多书籍都对实证主义范式做了详尽的介绍，《社会科学研究的思维要素》是其中最值得推荐的佳作之一：在25年中，能在美国先后出版9个版本（由于时间差，本书对照的是原书第8版）的学术性畅销著作，这显然意味着它在许多方面有过人之处。

这本小册子不是很深奥，事实上它是社会科学的入门读物，旨在帮助那些使用社会科学研究成果的人和那些将迈开研究生涯第一步的研究者。本书将处理诸如此类的基本问题：

- 概念从何而来?
- 何为变量?
- 为何要科学思维?
- 有关现实的假说如何区别于其他的陈述？等等。

我们力图帮助读者识破有关社会科学的某些错误印象，尽可能帮助他们踏出研究的第一步。全书强调的是现实检验，该检验是我们赖以认识世界构成的一种过程。这是对科学的一种泛泛介绍——我们鼓励读者既在日常思维中，也在社会科学方法的具体运用中保持科学的态度。



万卷方法博客圈:
http://q.blog.sina.com.cn/fafang
ISBN 978-7-5624-4465-7



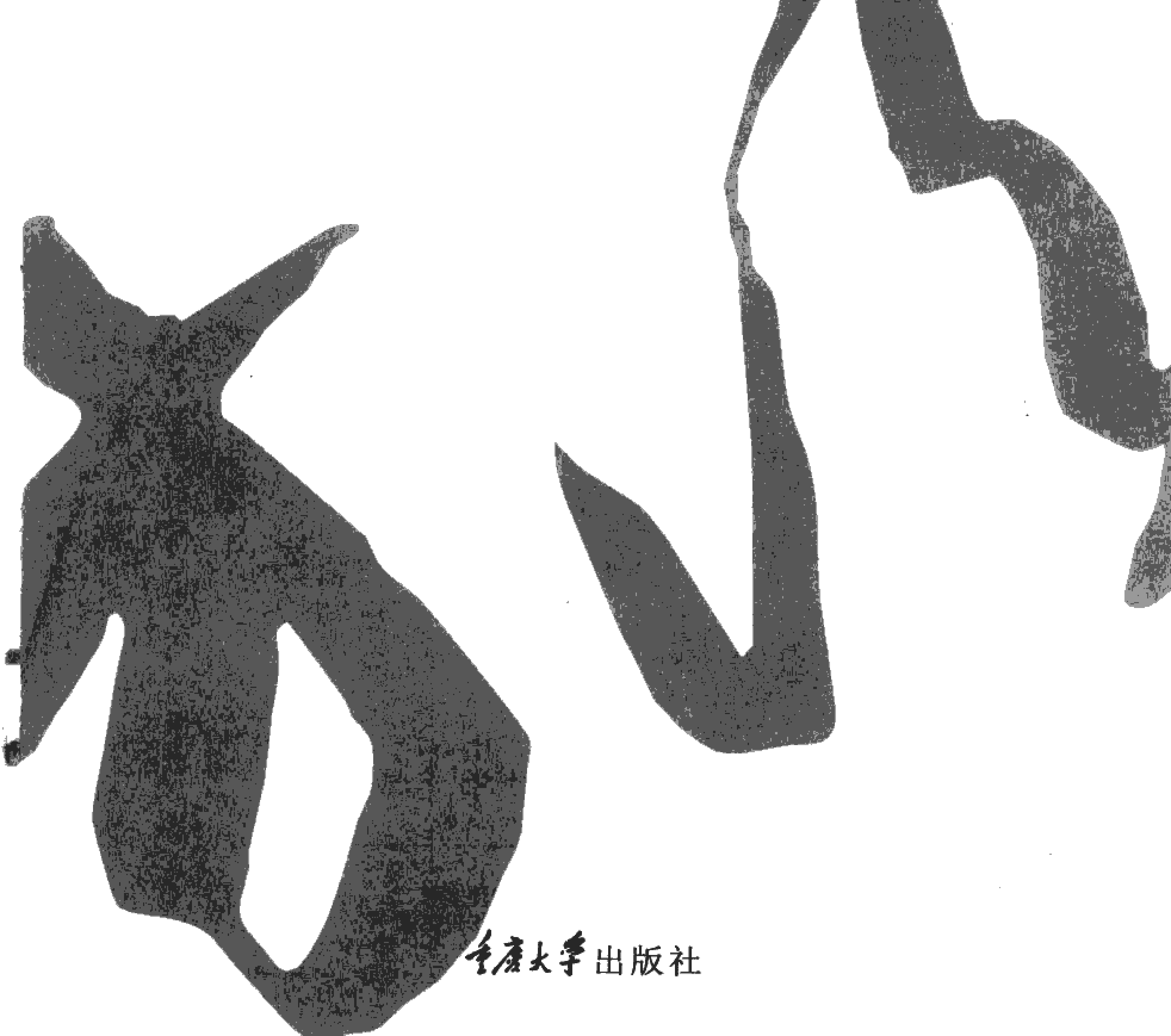
定价:25.00元

SHEHUI KEXUE
YANJIU
DE
SIWEI YAOSU

万卷方法

社会科学研究的思维要素

■ 肯尼斯·赫文　托德·多纳 著
■ 李涤非　潘 磊 译



重庆大学出版社

0-534-61411-6

图书在版编目(CIP)数据

社会科学研究的思维要素/(美)赫文(Hoover,K.),(美)多纳(Donovan,T.)著;李涤非,潘磊译.—重庆:重庆大学出版社,2008.4
(万卷方法)
书名原文:The Elements of Sociol Scientific Thingking
ISBN 978-7-5624-4465-7

Ⅰ.社… Ⅱ.①赫…②多…③李…④潘… Ⅲ.社会科学—研究方法 Ⅳ.C3

中国版本图书馆 CIP 数据核字(2008)第033828号

社会科学研究的思维要素
肯尼斯·赫文 托德·多纳 著
李涤非 潘 磊 译
责任编辑:雷少波 版式设计:雷少波
责任校对:任卓惠 责任印制:赵 晟
*
重庆大学出版社出版发行
出版人:张鸽盛
社址:重庆市沙坪坝正街174号重庆大学(A区)内
邮编:400030
电话:(023) 65102378 65105781
传真:(023) 65103686 65105565
网址:http://www.cqup.com.cn
邮箱:fxk@cqup.com.cn(市场营销部)
全国新华书店经销
自贡新华印刷厂印刷
*
开本:787×1092 1/16 印张:9 字数:166千
2008年4月第1版 2008年4月第1次印刷
印数:1—4 000
ISBN 978-7-5624-4465-7 定价:25.00元

作译者简介

肯尼斯·赫文 威斯康星大学的哲学博士，西部华盛顿大学政治科学教授。最近出版的著作有：《作为意识形态的经济学：凯恩斯、拉斯基·哈耶克与当代政治学的风起》(*Economics as Ideology*: *Keynes*, *Laski Hayek and the Creation of Contemporary Politics*, Rowman Littlefield Lanham, MD: 2003)。他还与 John Miles, Vernon Johnson, Sara Weir 合著《意识形态与政治生活》(第三版)(*Ideology and Political Life*, 3rd *ed.*, *Wadsworth*, 2001)。

托德·多纳 加州大学湖滨分校的哲学博士，西部华盛顿大学政治科学教授。最近出版的作品有：与 Shaun Bowler 合著的《共和国改革：新美国的选举制度》(*Reforming the Republic*: *Electoral Institutions for the New America*, 2004)，与 Shaun Bowler, David Brockington 合著的《选举改革与少数族群的代表性：排他选举的局部试验》(*Electoral Reform and Minority Representation*: *Local Experiments with Alternative Elections*, 2003)。

李涤非 武汉大学哲学博士，河南财经学院讲师。在《自然辩证法通讯》、《自然辩证法研究》、《世界哲学》等期刊上发表十多篇论文、译文，独译《做自然主义研究——方法指南》(重庆大学出版社，2007年9月)。

潘　磊 武汉大学哲学博士，武汉大学哲学系讲师。在《武汉大学学报》、《科学、技术与辩证法》等核心期刊上发表数篇论文。

译者前言

尽管最近几十年来不断有人著书立说，试图为社会科学研究树立一种“建构主义”（亦称“自然主义”）模式，以取代盛行的实证主义范式。无论取代的理由如何，但有一件几乎能被所有人认同的事实是，实证主义的研究成果正在越来越广泛地影响到社会生活的方方面面，并且在许多领域、尤其在经济领域取得了巨大的成就，为人们预测乃至控制社会现象朝良性发展做出了巨大贡献；相较起来，预测、控制社会现象从来就不是建构主义范式的主旨，要取得对等的成就即便不是完全不可能，似乎也是一件遥远的事。因此，实证主义范式仍是社会科学研究的首选模式，任何人想要迈入社会科学研究领域，都应该通晓这种范式的特点、构成和研究步骤。

有很多书籍都对实证主义范式做了详尽的介绍，但《社会科学研究的思维要素》是最值得推荐的佳作之一：在25年中，能在美国出到第9版（由于时间差，本书对照的是原书第8版）的学术性畅销著作，显然意味着它在许多方面的过人之处。本书尽管篇幅不大，内容也不像一般的著作那样深奥，但它有力地论证了科学思维方式的重要性，简练清晰地介绍了科学思维的各种要素，包括概念、变量、假说等，还用典型的研究范例展示了社会科学问题的研究步骤。当然，这里的介绍有点画蛇添足之嫌，因为作者本人在前言中就有比较具体的介绍，而再具体一些的就是著作本身了。我们建议，无论是已经在从事社会科学研究的学者，还是即将踏入社会科学门槛的新生，都应好好地阅读这本《社会科学研究的思维要素》；该书对于前者也许有启发作用，对于后者则是优秀的入门教材。

原书的译文是合作努力的结果，由李涤非和潘磊协力完成，分工如下：李涤非翻译前言和第1～4章，潘磊翻译第5章、第6章和附录A、附录B，最后由李涤非统稿。为了保证翻译质量，我们在校对时参考了台湾张家麟博士（在此对张博士致以谢意）的译本，在个别地方吸取了其精华（如译文第50页，“成为一位有远见卓识却昙花一现的女英雄”中的“昙花一现”），不过也避免了出现在张译本中的一些不足甚至错误。当然，不论我们做到如何的谨慎细致，总有不如意的地方，还恳请读者批评指正。

最后特别感谢重庆大学出版社的雷少波和陈进对我们的厚爱，把如此一本重要著作的翻译工作托付给我们，希望我们的努力对得起他们的信任。

译　者

2008年1月

作者前言

这本小册子不是很深奥，事实上它是社会科学的入门读物，旨在帮助那些使用社会科学研究成果的人和那些将迈开研究生涯第一步的研究者。概念从何而来？何为变量？为何要科学思维？有关现实的假说如何区别于其他的陈述？其相似性是什么？本书将处理诸如此类的基本问题。

我们的意图在于帮助读者识破有关社会科学的某些错误印象，尽可能帮助他们踏出研究的第一步，至于研究技巧的详尽细节，则有待参考更详细、更专业的资料。全书强调的是现实检验，该检验是我们赖以认识世界构成的一种过程。这是对科学的一种泛泛介绍——我们鼓励读者既在日常思维中，也在社会科学方法的具体运用中保持科学的态度。

第 8 版的变化

这本书最初出版于 25 年前，当时作者肯尼斯·赫文(Kenneth Hoover)还是一位年轻的政治理论学者，他回顾了自己在社会科学方面所受的教育，尝试着厘清一些关键概念和技术，以便新一代的大学生能够理解它们。作者最初反感定量分析和政治学的科学研究方法。不过，他逐渐认识到，它们有助于回答汉拿·皮特金(Hanna Pitkin)认定的作为理论家都要涉及的根本问题：如何改善人类的境况？我们无力改变的是什么？或许本书的长寿归功于它起源于方法论之外的领域，归功于它有意帮助那些对社会变化感兴趣的学生掌握社会科学的一些工具。

在第 8 版《社会科学研究的思维要素》中，我们延续了理论家肯尼斯·赫文和行为分析家托德·多纳(Todd Donovan)的合作，他们完成了此次修订的主要工作。多纳运用了大量的调查和统计技术，集中关注公民如何能参与大众民主。我们一直致力于以其他形式的“认知”为参考来定位科学知识，以便学生能识破阻碍有关讨论的某些陈腔滥调。核心教导依然是：通过观察我们只能局部认识构成世界的各种现象，而科学旨在减少世界的不确定性。此书的重点在于了解观察能做到什么。

我们还一贯强调直接的说明。我们做了许多小改动，然而，统计工具日益增长的复杂性，以及有助于这些工具传播的计算机的广泛普及，使我们觉得有必要扩展与研究技术相关的内容。为了满足社会学家、政治学者，以及其他社会科学家的广泛需要，我们改换了一个附录，以收入一些研究问题的范例。新的附录 A 介绍了有关美国“社会资本”衰退的讨论，以表明学者是如何研究社区参与和政治参与之间

逐渐变化着的关系。

我们确信,在那些做定量研究和不做这种研究的社会科学家(以及所谓的实证主义者和反实证主义者)之间的争论,有助于拓宽社会科学家的研究工具和研究前景。这种争论也消耗了大量能量和精力,现在最好是善用它们,使我们能对社会分析技术做建设性的使用。不说其他,显而易见的是,没有一种方法可以提供所有答案,任何方法都有其独特的缺陷,都可能导致偏见。选择合适的方法论或方法论的组合,是关键的考虑。

当社会科学家投入这些争论时,社会问题似乎变得愈加复杂、更难解决。如果审慎观察对于理解这些问题至关重要,那么社会科学就扮演了一个关键角色。

经典的科学研究规则提供了一种框架,用于解决在真理这个最有争议的主题上的冲突,即便冲突发生于彼此特别反感的人之间。我们尤其需要用系统的分析来考察源自各种洞见的有用观念,以便能把彼此冲突的观点消解在富有成果的活动形式之中。

如何阅读本书

大多数书籍适合于从头到尾一气贯通地阅读。对于许多读者而言,这也是阅读本书的最好方法。然而,读者应该意识到,本书的各章对社会科学思维的考察乃是立足于不同的层次。基于该原因,本书的入手有各种出发点,依读者的需要而定。人们对有关周边世界的问题的探索性回答有一些方法,第 1 章“科学化的思维”就是把社会科学置于这样一个普遍的语境中。第 2 章“科学的要素”通过讨论概念、变量、测量、假说和理论,提出了科学方法的基本轮廓。

对于那些直接承担研究任务或理解研究的人来说,第 3 章“策略”是个很好的起点,因为它直接讨论了科学研究的具体细节。第 4 章“提炼”假定读者对第 2 章中说明的科学方法已经有了基本理解,在此基础上提供了额外的研究工具。第 5 章“变量与关系的测量”探讨的是测量的艺术和科学。我们认为,出于任何目的而使用本书的人都应该阅读第 6 章“反思:追根溯源”。该结论章的要旨是摆正科学理解的位置,并从总体上指出,在哪里应该做到谦卑,在哪里可能取得成就。

为了方便读者温习,在每章结尾,依出现的先后顺序编排了该章的主要概念。

附录 A 摘选了罗伯特·普特南(Robert Putnam)所写的一篇题为《调来调去:美国社会资本的奇怪消失》(*Tuning in and Tuning out*:*The Ftrange Disappearance of Social Capital in America*)的文章。本书频繁引用了该文,那些需要一个好的模型来生成研究问题的人应该细心阅读。附录 B 是由大卫·布洛金顿(David Brockington)、托德·多纳(Todd Donovan)、肖恩·鲍勒(Shaun Bowler)和罗伯特·布里斯切托(Robert Brischetto)合写的一篇文章《累积和限制投票制下的少数族群的代表性》(*Minority Representation under Cumulative and Limited Voting*),该附录的内容仅仅与第 5 章中的回归分析那一节有关。

欢迎读者对本书提出意见,联系方式:Department of Political Science, Western Washington University, Bellingham, WA 98225,或者通过电子邮箱:Ken. Hoove@wwu. edu 或donovan@cc. wwu. edu。

目录

第1章 科学化的思维

Thinking Scientifically

“科学探索人们的共同经验；它被人建构，具有人的风格。”

——雅各布·布鲁诺维奇(Jacob Bronowski)

“社会科学”这个冷冰冰的字眼给人的印象是，一些机器人在统计实验室里把人类行为化归为毫无生气的数字和简单的公式。研究报告充斥着诸如“经验的”、“定量的”、“操作的”、“反向的”和“相关的”这些生硬的词语，没什么诗意。但我们将力图表明，对社会科学的这种老套看法是错误的。

与其他的认知模式一样，社会科学可以用于邪恶的目的，但也可以用来促进仁慈之士的个人理解。利用对现实的观察来检验思想，科学有助于解放研究，使之免于成见、偏见和十足的混乱。因此，仅仅由于老套的看法而退缩不前，不是明智的做法：太多人接受了科学的老套形象，因而把他们自己从事社会科学理解的能力拒之门外。

在我们的文化中，科学一词涵盖了许多意义，有些人甚至认为它是宗教在现代的继承者。在这里，我们的目的不是考察有关科学的所有纠葛，而是要找到一条通向科学思维的途径。为此，我们首先对比其他形式的知识，来看看对科学的某些描述。

首先，我们需要识别一些应加以忽略的令人分心之事。科学有时被混同于技术，技术是科学在各种任务上的应用。小学课本在太空遨游图片旁附上“科学在前进！”的标题，助长了这种混淆。技术使太空遨游成为可能，它使用了推进力的研究、电子学和许多其他领域里的科学策略。探索模型才是科学的；太空船则从属于技术。

正如科学不同于技术，科学也不是某种具体的知识体系。“科学告诉了我们(比如)吸烟可致命”，这句流行的话实际上是一种误导。“科学”不告诉我们任何事情；告诉我们事情的是人——在这个例子中，是那些运用科学策略调查了吸烟与癌症之间关系的人。科学作为一种思维和调查方法，我们不应

把它看做存在于书本、机械和含有数字的报告中，而最好认为它存在于心灵这个无形的世界中。科学与问题的提出和回答方式相关；它是用于探索的一套规则和形式，由那些寻找可靠答案的人们创造。

另一个分心之事是把某些特殊的人认作“科学家”。这种用法并不错，因为被这样称呼的人，实践的是科学的探索形式；但说某些人是科学家，而其他人不是，并不是完全公正的做法。虽然有些人精于用科学方法获取知识，但我们所有人都是科学思维方法的实践者。科学是一种探索模型，为全人类共同拥有。

当你更深刻地意识到自己的思维习惯时，你就会发现，我们每个人都有那么点“科学家”的意味。我们测量、比较和调整信念，获得对日常事务中的有关证据的理解，为下一步计划做出打算，并找出与别人打交道的方法。最简单的游戏也牵涉到运用执行数据来检验方法和策略，而这已初具科学的形态了。即便为了给他人留下深刻印象而试穿不同风格的服饰，也包含科学的要素。

我们有许多策略用于应付重大的现实问题——生命的不确定性，科学思维方法只是其中的一种。我们并不知道许多行为的结果。我们也可能不清楚某些压力，它们在微妙地或直接地、逐渐地或陡然地影响着我们。即便在尝试完成最简单的任务时，比如计划吃什么，我们也在做一些基本的盘算：哪些食物的味道好，哪些食物对我们有益。如果这样做还是太不确定，那么稍微深入的检验是个不错的主意：女王有品尝员，而我们其余的人至少在看到某种汉堡时，也可以确定它早就卖出数十亿个了。

科学是一种思考和提出问题的过程，而非一种知识体系。我们有许多方式声称我们知道某事，科学只是其中的一种。在某种意义上，科学方法是一套标准，用于决定如何解决不同的现实观之间的冲突。科学为研究者提供问题研究的策略。科学为运用研究结果的人提供一种能力，使他们能批判性地评估如何提出和使用证据来达到结论。

在寻求理解方面，科学方法有许多竞争者。纵观绝大部分历史，对于许多人来说，获胜的是那些竞争者。对现实的分析一般不如神秘、迷信和预感那样流行，后者在它们试图预测或控制的事件发生之前使人有确定感，尽管事后发现很少如此。有时候，未被证实的信念促发一种受激发的行为，或是让怀疑者继续犹豫不前，直到更好的时机来临。确实，某些个人信念构成了我们生命的重要部分。关键是，拒绝分析是自断其臂，而熟练的分析者则占得先机。

为何要系统化

大多数人类交流发生在小群体中，他们拥有共同的语言、大量的对周边世

界的共同经验和理解，要达成彼此间的一致很容易。但在较为复杂的社会环境中则很难如此。尽管家庭能通过故事和箴言的传承，在代与代之间传递智慧，但在社会中却有麻烦。用最悲观的形式来表达，问题就是："要相信谁的故事？"理解周边世界、与其他人分享经验的需要，使系统的思维和探索变得至关重要。

社会因其戏剧性的事件而引人入胜，因此人们倾向于摒弃系统的理解，而习惯于描述、故事和个人判断。尽管它们可能具有启发性，但通常作用有限，因为对生活所做的高度主观的描绘，对于发展共同理解和共同活动来说，是糟糕的基础。

要在各自经验的独特性而造成的差异之间架起沟通的桥梁，这种复杂的任务需要运用训练有素的方式来获取知识。只有能加以运用的知识，才能成为推动社会发展的力量。社会知识要有用，必须是能交流的、有效的和有说服力的。

为了能被交流，知识的表达形式必须清晰。如果要把知识用于促发行为，它必须是有效的，这就要求有恰当的证据；还必须是有说服力的；它要适合于被提出的问题。"我认为资本主义在剥削穷人"，这类个人意见可能令你的朋友，甚至亲戚认为，我们的社会存在某种不公。然而，如果你能举证，在美国六个孩子中至少有一个生活在贫困中，你就会得出更有说服力的结论，因为你把判断与一种对现实的测量关联起来了。[①] 那些即便不喜欢你但提倡财富分配公正的人，可能就会觉得这种陈述对于批判性地考察我们的经济体系，是条有力的线索。知识建立于证据之上，以清晰的、可传达的形式表述，促进对环境的改造。

积累知识以避免过去所犯的错误，通常会孕育出文明化的人性。人们能记录智者的格言，这为文化的发展作出了极大的贡献。但无可置疑，存在另一种积累方式：建立受证据支持且可以被他人加以复查的陈述。对陈述的复查，需要当事人准确地知道陈述的是什么，如何检验陈述。这是科学事业的一个主要组成部分。第2章关于科学方法的一节所讨论的步骤，是完成那类认知过程的指导方针。

---

① 参见 National Center for Children in Poverty, "Low Income Children in the United States: A Brief Demographic Profile." Mailman School of Public Health. Columbia University. April 4, 2002 (http://www.nccp.org)。该报告使用了美国人口普查机构从1976年3月到2001年3月的数据，而且指出，15.8%的儿童生活在贫困中。这是自1978年以来贫困儿童的最低水平。

有理判断和意见的角色

关于系统思维的所有这些带有模糊预示性的说法，并不是要排除有理判断、意见和想象。毕竟，在任何知识探索中限制心灵的能力，都不怎么明智。

有理判断是人类理解的主要组成部分。有理判断与证据之间有相当程度的关系。因为人们难免在缺乏完整证据以供决策的情况下采取行动，所以“判断”这个词语很重要。判断意味着决策，在该过程中，心灵的所有能力都被调动，以使可资利用的知识发挥最大效用。

社会科学并不排除判断在研究过程中的角色。事实上，在科学证据的搜集和评估中，判断扮演着重要的角色。我们可以观察到，美国前 10% 的家庭的收入比后 10% 的穷人的收入高 55 倍。最富有家庭的收入在最近的经济繁荣期以 15% 的速度递增，穷人的收入则停滞不前。不过，把该证据与资本主义、不平等、贫困、财富、剥削、生产力、经济发展以及其他广泛的社会问题关联起来，则是另一回事。我们需要用逻辑和良好的判断来解释这个证据。②

有理判断是系统思维的第一个部分。“选举日的前夕若是满月，则能促进自由派投票”，这个命题可能是正确的，但它并没有反映出多大的有理判断，因为既无证据把两个事件连接起来，它们之间也无逻辑上的关联。有空闲又有资源的研究者可以考察这样一个命题，但在这个世界中，时间紧张、资源匮乏、社会分析占用的都是稀缺人才这些严重问题充斥着的世界中，这样的研究可就没什么意义。③ 尽管那个命题可能合乎直觉，但即使是直觉也通常与经验及证据有某种关系。

意见在科学分析中同样扮演一个必不可少的角色，因为所有的探索努力，都源自某种个人兴趣。只有对结论感兴趣，人们才会提出问题。此外，不同的人观察现实的角度必然会略有不同。探索中的意见不能消除，但能加以控制，不致使其变成十足的幻想。对于研究者而言，有助于减少意见角色的一种方法是，留意自己的价值和意见。

---

② 参见 Shannon McCaffery, “Income Gap Growing.” Associated Press Wire Service (January 18, 2000)。Richard W. Stevenson, “Income Gap Widens Between Rich and Poor in 5 States and Narrows in 1.” *New York Times* (April 24, 2002). Michael Hout and Samuel Lucas, “Narrowing the Income Gap between Rich and Poor,” *Chronicle of Higher Education*, Section 2 (August 16, 1996), B1-B2。这些研究表明，在 1996 年，富人和穷人在收入上的差距自 1921 年后达到最大，而且在继续拉大。

③ 不过，警察和酒吧侍者会告诉你，事实上，满月之夜会发生相当奇怪的行为，这个假说并不完全是荒谬的。

这里用得上柏拉图的一句名言:"认识你自己。"有些人自诩他们是客观的,他们在研究中隐藏了自己的意见,而实际上正是这些意见悄然地建构了他们的结论,这些人对优秀的社会科学事业造成了大量伤害。没有人是真正客观的,对于社会本质的看法当然也不是客观的——其中掺杂了许多个人利益。

从根本上说,优秀的科学自身会考察探索中的价值所造成的影响。如果研究方法和用于支持结论的证据被清晰、充分地陈述,我们就能检查结论与证据的匹配度。如果对研究过程的有效性抱有怀疑,我们还能够对研究本身进行复查,或用技术化的术语来说,就是"复制"。这一特征把科学与个人判断区分开来,并使科学免于个人成见。

不像心灵处理内在的感受、对经验的知觉和思维进程,没有人能够复查所有发生的事情。科学把探索步骤带出心灵,引入大众的视野中,使之作为知识积累过程的组成部分而被所有人共享。

想象、直觉和习俗的角色

心灵有许多认知方式,但都不如在想象中显得那么巧妙却又那么神秘。如果人们还能够一跃跨过高高的障碍物的话,那就是心灵的飞跃。但是,想象一条关于现实的可能命题是一回事,而着手去想象证据,则完全是另一回事。

科学的实质在于找出我们有所了解的事物间的关系。提出一种关系,是一种创造性和想象性的活动,然而,这可能需要大量系统的准备工作。用现实检验命题,牵涉到不同层次的想象——主要是一种能力:从那些源于现实、零零散散的信息片断中,找到一条信息,它对于检验某个特定观念的可信性而言是必不可少的。

正是在探索发现领域,科学与想象结成了亲密伙伴。自然科学的历史中到处都是这类例子,既包括认识到地球围绕太阳运转,而不是相反,也包括发现物质是由微小的原子所构成的。其中每项发现都是由那些大胆而富有想象力的人做出的,他们参照现实世界中的证据,不怕挑战惯常的信念结构。尽管这些都是大尺度的发现,但都包含相同的努力:步出对人类行为所做的公认说明,想象其他的可能性,并合理地运用证据来检验它们。当女性主义者在考察男女区别的传统看法时,就是这么做的。要真正富有想象力,有点类似于试图摆脱重力的束缚——万事开头难。虽然到现在为止,社会科学中很少有能够与自然科学中的丰功伟绩相匹敌的发现,但把科学运用于社会关系,只是最近

的事,而且是复杂得多的工程。④

从基本层次上说,正是有了好奇心,想要发现无序中的有序,才有了科学探索。看到一系列令人困惑的事件、事变和行为,我们急于了解为什么某些事发生了,事件产生的原因是什么。社会科学能满足我们的好奇心,使我们获得理解。从另一层次来说,社会科学制造可以交流、可以用来把我们的理解解释给他人听的知识。

无论我们对科学分析所要求的审慎思考持什么样的看法,仍旧无法完全把握"有个主意"这一奇妙的过程。科学决不是排斥直觉和想象的一个体系,相反,它是一套程序,使这类观念在人类智力允许的范围内尽可能变得富有成果和创造性。即使是最奇妙的观念,其作用也来源于它与当前的或潜在的某种现实之间的关系。科学是一门现实检验艺术:形成观念并从有关现象中引出可观察的证据来检验它们。

步出一般模糊不清的人类关系,构想其他的可能性,需要一种异乎寻常而又必不可少的想象力。说来遗憾的是,就通常的社会和政治经验而言,大卫·休谟(David Hume)的观察或许是正确的:"(人们)一旦习惯顺从,就不会想到步出他们及其先祖踏出的道路,步出这条被许多急迫且明显动机所限制的道路。"⑤然而,正是在对社会、政治布局的理解和革新上,世界需要严谨想象的最佳运用。如果在缺乏想象的情况下理解社会现实,我们就会困在习俗这条老路上,就会陷入窒息人类潜能的不公之中。

同样,我们也许会局限于一些徒劳的行为习惯。英格兰曾经有个惯例:当众绞死扒手以震慑其他人的偷盗行为。然而,有人观察到,比起其他公开事件来说,越在绞刑的场合,偷盗事件越多。这种惯例存在的时间超过了这么点社会科学的发展历史,其实它远不该存在那么久的。

习俗并非全然是坏事,因为它可能包含长期以来的教训,学自对现实的体验,经常是不愉快的体验——在不严格的意义上,习俗算是科学的。习俗经常在大量的,甚至激烈的压力下把社会群体团结在一起。不过,任何社会科学的

④ 也许用科学考察社会习俗的最早努力之一,是英国科学家弗朗西斯·盖尔顿(Francis Galton)在19世纪末做出的,他检验的是祈祷的灵验。他观察到,每天在各地教堂里都有祈祷皇室长命百岁的祷告,他把他们的寿命与贵族们的和各种职业人士的寿命做了比较。他发现,排除意外或暴力导致的死亡,并只涵盖那些寿命超过30岁的人群,皇室成员的平均寿命是64.04岁,这在他的分类观察中是最低的。不过,盖尔顿确实注意到,除了满足一些需求外,祈祷还有许多个人用途。而且,谁又知道,如果不做这样的祷告,皇室成员可能更短寿呢。P. B. Medwar, *Induction and Intuition in Scientific Thought* (Philadelphia: American Philosophical Society, 1969), pp. 2-7.

⑤ David Hume, "Of the Origins of Government," *Political Essays*, ed. Charles Hendel (New York: Liberal Arts Press, 1953), p. 41.

任务都必须是理解事物的存在方式,理解如何革新社会生活的元素,以适合更人性化的个人发展与表达的模型。在追求理解时所使用的工具,不只是用训练有素的程序进行探索的科学,还包括一些直觉:生活能过得比现状更好,某种行为模型可能并非不可避免,微不足道的行为交流或许是向更大的可能性和潜能迈进的关键。

任何理解工作所采用的方法,都牵涉到思维与调查之间的张力。连接这两个要素的方式各种各样。神秘主义者察觉到一种内在的真理,用一些"征兆"来解释其洞见的有效性。历史学家从过去的历史中寻找模型,并认为它们可以用于解释事件的意义。比如,欧洲"中产阶级的崛起"成了历史学家用于解释的主要概念。而具有科学精神的人努力做到比神秘主义者更具体,比历史学家更精确,他们关注研究的指导思想,调查中被视为重要的数据,以及对照现实以检验心理建构的测量方法。

在后面几章,我们将逐一考察建构科学理解的步骤。正如你将看到的,这种技术更多的是需要常识,而不是技术知识或精心准备。

引入的概念

- 科学(science)
- 技术(technology)
- 可交流的知识(communicable knowledge)
- 有效知识(valid knowledge)
- 有说服力的知识(compelling knowledge)
- 有理判断(reasoned judgment)
- 意见(opinion)
- 客观性(objectivity)
- 想象(imagination)
- 直觉(intuition)
- 习俗(custom)

讨论

1. 非科学的理解模式有何实例?如何用这些非科学的模式来解释下述问题?
   - 为什么某些国家比另一些国家富有?
   - 为什么政治革命发生在某些地方而不是在其他地方?
   - 明年的世界职业棒球大赛中谁会获胜?
2. 社会科学知识如何比其他的知识形式(比如直觉、习俗)更有效力?其缺点和危险是什么?
3. 科学知识如何用于改革或改变社会?
4. 为何想象对于社会科学至关重要?
5. 想象的运用对于社会科学来说,是否比对自然科学(如化学、生物学)更重要?

第2章 科学的要素

The Elements of Science

“(科学探索)起始于我们发明的一个关于可能世界的故事,随着推进,我们批判和调整它,最终尽可能使之变成一个关于现实世界的故事。”

——P. B. 梅达沃(P. B. Medawar)

我们已经尝试着参照其他类型的思维讨论了科学思维,指出了我们为什么应该对科学感兴趣。现在该是我们了解科学构成的时候了。

科学策略的元素本身很容易理解。它们是一些概念、变量、假说、测量和理论。这些元素的组合方式,构成了科学方法。理论的功能是引导出这种方法、赋予它意义,这是通过帮助我们解释被观察到的现象来实现的。首先,我们试着摆正每个要素的位置。

概念的起源和功用

假如清除你心中所有的词语和其他符号,以未开化的心灵面对世界,你会做什么?没有身体的支撑,你可能什么都做不了。不过,生存的必要性毋须多谈,而且,心灵的最初行为就是区分能吃的和不能吃的,接着是区分冷热和敌我。这就离“食物”、“庇护所”和“温暖”这类概念的形成不远了,也离用词语或言辞来符号化这些概念不远了。这样就粗陋地出现了语言这个工具。对真正有用的概念和范畴的探索已经起步了。语言就是大量概念的集合,这些概念包括事物、感受以及观念的名称,它们是人们在彼此互动和与环境的互动中产生或习得的。

有些概念和分类可能不是很有用。只用一个名称来概念化所有的植物,就会阻碍对可食用的、有药用价值的以及有毒的植物的深入区分。有些概念与经验的关系太含糊:英语中只有一个词语表达“爱”这种外延广泛而复杂的

情感，而希腊语中却有三个概念：eros 表示浪漫的情爱，agape 表示一般化的慈爱感情，filios 表达的则是家庭之爱。英语在处理爱这个概念上的不足，通过该词语在我们文化中的不便运用，影响到每个人的经验。

注意，现实检验正好构成了事物命名过程，它是存在物最基本的事务。环境刺激与心灵沉思之间的这种互动，构成了我们试图掌握以用于分析的那类思维。

数千年后，我们仍然要面对一个事实：命名是一个困难的过程。本质上，语言源于约定。例如，家庭（部落、民族、国家和世界）中的你、我和其他成员约定，把天空中闪烁的东西称为星星。遗憾的是，这些约定可能不是很精确。就通常的用法来看，词语“星星”涵盖大量的对象，有大有小、有热有冷、有固态也有气态。

用精确的名字称呼事物，是理解的开始，因为它是心灵把握现实及其众多关系的关键。认为生病是由于糟糕的精神状态引起，还是由于放纵时被细菌感染所致，其中有重大的区别。“细菌”这个概念依附于一个概念系统，后者又关联到一个有效的治疗体系，即抗生学。

即便草率一点，要把握语言的意义，也是一个深奥、微妙的过程。例如，“种族”这个抽象的概念表达的是团体身份的差异。当赋予种族范畴或属性以名称时，命名的问题、影响力和困难性就显而易见了。研究者经常采用简单的种族分类，用“白人”和“非白人”（或“盎格鲁的”和“非盎格鲁的”）来给人命名。这种划分尽管是通行的做法，但抹杀了世上大部分人之间的区别。此外，名称本身也能引起复杂的问题。想想在美国有各式各样的名字用于指称美洲黑人（黑鬼、美洲黑人、有色人种、黑人）或美籍西班牙人（拉丁美洲人、拉丁美洲女性、美籍西班牙人、有色人种、拉丁美洲人、墨西哥美洲人、中美洲人、波多黎各人，等等）。

命名这个过程也能给命名人带来重大的影响力。种族这个概念的特性不容易被命名。此外，种族名称是对不同人的不同定位。你自己对社会科学思维的表达，尽管要求你要精确对待概念，但不是要你对你的新知识在改变生活、社会和文明方面的功用太过自信。

恰当命名事物的重要性，再怎么强调也不为过。17 世纪的政治学家托马斯·霍布斯（Thomas Hobbes）认为，恰当命名事物对于政治秩序的建立至关重要，他甚至认为这是君主的一个核心职责。詹姆士国王知道个中要害，命人对《圣经》做一种官方权威的诠释，以平息那些在圣经文句用词的精确意义上的激烈争论。

与现代关联更紧密的是下面这个例子：乔治·奥维尔（George Orwell）在他的反乌托邦小说《1984》中描绘了一幅景象：一整套官僚机构投身于语言概念的重构中，以此提升极权主义社会的影响力。克林顿总统曾试图通过重新

定义自愿成年男女间的性关系这一概念,来消除有关他与莫妮卡·莱温斯基(Monical Lewinsky)之关系的争议。这些事例旨在让你意识到,对概念意义的修改,可以影响到人类理解和社会控制的基础。

然而,在你掌握科学方法、足以让你建立丰功伟绩之前,还得等段时间。眼下的关键是,对于科学目的来说,概念是:(1)尝试性的,(2)基于共识的,(3)其作用只能局限于把握或分离现实中的某些重要的、可定义的方面。

概念与科学有何关系?如果你在婴儿身边呆过一阵子,你可能注意到,他们经常试着指向各种物体并给它们命名来显摆。尽管这个动作会重复10到15遍,有点烦人,但他们对这种练习感到自豪,而且这是有道理的。下一步就是造句了。首先是命名事物,符号化它们而不是仅仅指着它们,接着就是靠近现实,从而能制造他们需要的事物。安德鲁·赫文(Andrew Hoover)的第一句话是对他姐姐艾黎说的。他坐在小手推车里说:"艾黎,推推!"艾黎照做了。

你现在看到的内容是一种努力,它试图把概念联结起来,以拓展你的理解。人们说成千上万的话,只是为了使现实做出某种有益的回应。大多数人不像安德鲁那么幸运,第一次开口说话就能产生效果。通常而言,概念容易混淆,关联也很模糊或不可靠,更不用说讲话人可能不清楚听者的理解力和动机这个问题了。

通过联结概念,思维和理论得以发展。考虑一下皮尔·普鲁东(Pierre Proudhon)的名言"财产就是偷窃!"。作为一个概念,财产表示一种观念:个人可以声称土地或其他资源为他所独占。当然,偷窃意味着不正当地攫取某物。普鲁东用动词"是"把两个概念联结起来,旨在传达这样的一个信息:私人财产制度等于否认人类对自然资源的共同所有权。他认为,私人专有财产是不正当的偷窃行为。尽管普鲁东是从哲学的高度揭示了两个概念的关联,但那句最朴实无华的话就可以实现相同的功能。

科学是一种方法,通过参考可观察的现象来检验概念的表达以及它们之间的可能关联。我们下一步就来看看科学家是如何把概念转变为可观察之物的。当概念被定义为变量时,就能用于构造一类特殊的语句,即假说。

何为变量

变量是一物之名,能对其他某物的一种具体状态施加影响,或者是被后者影响。热量和压力都是可以使水沸腾的变量。年龄在投票中是一个较为重要的变量,不过,还有许多其他更加重要的变量:社会经济地位、父母的影响力、种族、性别、居住地,等等。

此外,变量还是一类特殊的概念,其中蕴含着对程度或差异度的看法。温

度就是一个比较容易理解的例子，它包含了对热量多少的看法，也就是说，对程度的看法。顾名思义，变量指称的是变化着的事物。社会科学感兴趣的焦点是那些牵涉变化的概念，是探讨一个现象的变化何以能说明另一现象的变化。

举个例子，考虑一下宗教与投票的关系。首先，宗教不同于温度这样的变量。尽管可能存在“虔诚”度这回事，[①]但我们很可能是根据宗教派别来谈论宗教这个概念中所包含的变化。人们认同的宗教存在着实质的变化。例如，在2000年总统大选时，发布的民意调查数据被用于评估宗教在选举中的重要性，调查显示，民主党总统候选人阿尔·戈尔（Al Gore）以48.4%的支持率超过了共和党乔治·布什（George W. Bush）的47.9%和绿党候选人拉尔夫·纳德（Ralph Nader）的2.7%。布什赢得了选举团的投票，在佛罗里达最高法院终止引起争议的重新计票后，他当选了总统。《洛杉矶时报》收集的数据显示，整整有80%的犹太选民支持戈尔（对比而言，布什为17%，纳德为1%），而63%的新教徒投给了布什（相比之下，戈尔为34%，纳德为2%）。52%的天主教徒投给了布什，而只有45%投给戈尔。“其他”宗教的信仰者以及声称自己无宗教偏向的人中，大多数把票投给了戈尔。[②] 诸如这样的数据有助于我们对变量宗教与变量投票行为之间的关系做出有意义的评价。

尽管大多数变量处理的是程度的差异，如温度的差别，或处理多样性的差异，比如宗教，但有些变量更加简单，它们处理的是最基本的变化：在场或不在场、有或没有、存在或不存在。例如怀孕，不存在多少的问题。要么有，要么没有。

把概念转换成变量，尽管看上去似乎枯燥无味，却是一个极富创造性的过程，而且经常引起有趣的问题。举个例子，考虑一下“时间”这个普通的变量。早期希腊人为概念化这个变量而绞尽脑汁。似乎显而易见的是，时间应该有个开端，因此哲学家力图推算出开端起于何时。然而总是跑出一个困惑的问题：在这之前发生了什么？

柏拉图和亚里士多德都曾有过这样的念头：时间可能根本不是线性的，也就是说，它可能没有开端，没有前进，也可能没有终点。它可能是周期性的！对于习惯线性时间观念的我们来说，这种看法似乎很疯狂，但他们两人想的是，宇宙时间可能有点像生理周期——在自然界随处可见的规则变化。因此，

---

① 根据上教堂的频率这样的态度和行为表现，人们进行了各种尝试来测量个人的“虔诚”度。例如，见 Lyman A. Kellstedt & Mark A. Noll, “Religion, Voting for President and Party Identification, 1948-1984,” *Religion and American Politics: From the Colonial Period to the 1980s*, ed. Mark A. Noll (Oxford, England: Oxford University Press, 1990), p. 347。

② CNN. com 2000 年选举报告。“Exit Polls.” http://www.cnn.com/election/2000/results/index.epolls.html，2002 年 12 月 2 日。

我们最好是把历时的时间看做是事件的演变结构，事件在其中依次发生，直到整个模式的结束，然后开始下一轮的循环。亚里士多德评论说，他自己很可能“生活在特洛伊城陷落之前，也在陷落之后，因为，当命运的车轮开始另一轮循环时，特洛伊战争将会重演，而特洛伊城会再次陷落。”③

刚入门的学生在从事社会科学研究时，很少卷入这类烦心的概念问题，但并不能假装这些问题不存在。例如，变量“人格”，据说在专业文献中有四百多种定义，部分是因为人格是一个复合变量，包含大量其他的变量：阶级、地位、自我概念、种族、社会化，等等。人格这一变量的复杂性迫使社会科学家采用了这样一个尴尬的定义：“人们在后天习得的、相对稳定但又变动的、心理和社会行为倾向的独特系统。”④

即便社会科学家都认同对某个变量的表述，也不意味着该定义是永恒的真理，而只是意味着，谨慎思考过它的一些人都认为，这个定义似乎有助于回答某些问题。此外，研究者常常为了方便而决定采纳对某个变量的一种定义。在美国，按照惯例，党派认同是通过人们对调查问题的回答来测量的，这些回答把选民置于一个连续统中，显示选民对于两个主要政党的认同度。该连续统包含 7 个类别：

←强民主党人士—弱民主党人士—偏民主党的无党派人士—

无党派人士—偏共和党的无党派人士—弱共和党人士—强共和党人士→

这里假定政治无党派人士居于政治连续统的中心。但真实情况可能是，许多“无党派人士”认为自己是激进分子，并不处在中心。有些人可能完全无党派倾向或不关心政治，因此他们认为自己压根就不应该用政党来归类。另外，某些“有偏向的”无党派人士从投票行为来看，与“弱”党人几乎一样有党派性。⑤尽管对变量的这一定义不可能完美地反映党性这个概念背后的实情，但它仍有预测力。相关问题已被调查了数十年，因此研究者可以评估一段时间内在党性问题上的趋势。由于在这个连续统上对无党派人士进行范畴化，其困难已经变得明显，因此出现了对党性的新定义。我们应该详细地阐述如何把概念转换成变量，忽视这个问题是掩耳盗铃的做法，只会使我们深陷语言的泥沼之中。

---

③ Stephen Toulmin & June Goodfield, *The Discovery of Time* (New York: Harper and Row, 1965), p. 46.

④ Gordon DiRenzo, *Personality and Politics* (Garden City, N. Y.: Anchor Books, 1974), p. 16.

⑤ William Flanigan & Nancy Zingale, *Political Behavior of the American Electorate*, 10th ed. (Washington, D. C.: Congressional Quarterly Press, 2002), pp. 85-87. 这些作者指出，随着越来越多的无党派人士甚至不敢自称独立，美国选民正变得“更为中立、但并非都变得更加独立”。

语言中库存了丰富的概念,为我们联结变量来解释事件提供了大量的可能性。数百年来,人们一直手忙脚乱地梳理有意义的关联。科学是一种稍微高级的处理形式,它试验这些关联,尽可能仔细地检验它们。在医学中,人们花了几个世纪才分离出许多影响疾病的变量。直到最近医学才成为一门成熟的学科,能利用极为有效的血液化学分析来诊断许多疾病。这种发展反映了其成熟经历了一个漫长的过程,其中要分离和消除大量不重要的或不那么重要的变量。西医越来越多地接触到一种在中国被发展成具有很高艺术造诣的古老医学体系,所以现在我们的医学家们力图弄清针灸为什么可以治病。我们必须考虑一套全新的变量,建立新的概念关联,克服习以为常的理解所造成的阻力。

遗憾的是,就社会科学来看,我们很少弄清该如何奠定理论结构的基础来解释社会行为。许多社会科学的新生看不出——尤其在面对厚厚的入门课程教材时,社会理解所取得的成就背后蕴藏着的努力与成绩、尝试和可能。

目前,社会科学包含许多分支(例如,政治学、社会学、经济学、心理学、教育学),所有这些学科都在忙于定义、观察和关联行为子系统中的特殊变量。社会科学家正在探索大量可能的变量关联。正在出现的各种已被检验的零碎知识,等待着跨越这些探索界线的整合。由于我们迫切需要全面的社会理解,因此这些努力必定会越来越多,但有关尝试还相对较少。

量化和测量:转概念为变量

我们在前面提过,社会科学家把概念转换为变量。这样做是为了用一种可观察的形式把概念表达出来,使之包含对程度或差异的某种看法。接下来的问题是:如何确定程度或差异?答案分为两个步骤:量化和测量。

量化的观念意味着为某个事物设定标准量,并贴上标记。用这种方式表达抽象的概念(如长度),有可能为观察提供共同的参考。有些量化的起源相当奇怪。例如,古希腊人需要一种衡量距离的标准,他们就以赫拉克勒斯的足长为长度标准。这只足与腕尺竞争长度标准,后者是某个人的前臂长。腕尺的问题是,人们根本不可能就标准的前臂长度达成一致——有人说是17英寸,有人却说是21英寸。结果,人们再没谈到腕尺这个长度标准。

分离标准化的单位,有助于描述和分析。当加布里埃尔·华伦海特(Gabriel Fahrenheit)形成了温度高低的观念后,他就能对热和冷做更有用的描述。如果水温是32度而不是33度,其中就存在巨大的差别;而词语“冷”和“更冷”是没办法很好地把握那一重大差别的。

在社会科学中量化有两种形式:不连续的和连续的。不连续量化包括计算事物的在场与不在场,也包括计算属性的差异,其前提是差异以范畴形式出现。投票支持某候选人是不连续的、特殊的行为,可以用常规的方式计算。个

体的性别是一种属性,可以算做是男性或女性。

不过,有些量化必须把握连续变异的观念。年龄是一个连续量化的例子。我们确实可以用年数计算人的年龄,但年龄的量化,表达的是演化着的东西。今天,本书的一位作者是62.29岁,明年9月他将63岁。连续量化处理的不是离散的事项,而是处理年龄、长度和时间这样的纬度。连续量化的标志是,有关变量可能在一数值范围内取任何值,而在不连续量化中,只能出现整数值(如计算羊的数目)。

每个变量的量化都有其独特的问题和潜能。成熟科学的显著特点之一是有一系列可量化的变量,能为有关领域的工作人员提供帮助。有天赋的科学家的一个特征是,他们有能力找到方法,以一种可靠的和有意义的方式量化那些重要的变量。经济学用金钱作为分析单位,走过了漫长历史(虽然除了其他人,甚至经济学家有时也会混淆金钱和价值)。许多有效的经济指标,如国内生产总值或消费价格指数,都建基在金钱之上。

遗憾的是,其他的社会科学没有这么容易量化的单位,来测量权力或表征心理压力、异化、幸福、个人安全,还有价值。不过,有创造力的科学家找到了一些多少有效的方法,把握到这些变量的一些可量化的方面。这些领域里的任何一本教科书中都包含数十种例证,表明如何把概念转换成可量化的变量,我们在下一章做一些介绍。量化的重要性在于,一旦能够被量化,我们就有可能做更精确的测量。

测量不是我们能选择的事,它内在于各种分析性的讨论中。如果你持怀疑的话,那在下次交谈的时候聚精会神地听听,看看你对那些包含测量要素的词语的依赖性。"民主党人一般偏向穷人",这样一个简单的政治陈述都包含三种测量。动词"偏向"意味着程度上的差异,"穷人"一词也是如此,"民主党人"是一种分类。修饰词"一般"试图限定这种测量,指出这并不是所有民主党人的一个普遍特征。

如果数量能被确定,测量就会容易得多。最明显的测量处理的是多少的问题:有多远、多少钱,等等。有些牵涉到多少的问题不大容易测量,比如民意。在把对问题的回答当做可量化的分析单位时,粗糙的调查方法给回答者提供的只是"被迫选择",把意见分为赞成或反对。在这里,意见是作为不连续的、绝对的变量被量化的:"你是赞成还是反对?"民意调查通常是以此为基础的。这么简单的测量无疑会掩盖民意的强烈程度。在许多政治问题上,可能有少数人在一种立场上持强烈的态度,而多数人不冷不热地持另一种立场。一些民意调查在处理这种情况时,使用了四个范畴而不是两个:

强烈支持　支持　反对　强烈反对

只以多数人的情感作为行动基础而不考虑强烈程度的政治体系,会给自身带来许多麻烦,就如美国在越南战争时的做法一样。态度强烈的少数反对

表 2.1 对总统和副总统候选人的感受的测量,2000

	平均评级	评级为“50”的百分比	无法归类的百分比
Al Gore, Jr(阿尔·戈尔)	57	15	2
George W. Bush(乔治·W·布什)	56	18	3
John McCain(乔·麦坎)	60	25	21
Dick Cheney(迪克·程伲)	56	28	24
Ralph Nader(雷普·纳德)	52	30	27
Joseph Lieberman(约瑟夫·利普曼)	57	28	27
Pat Buchanan(帕特·布坎南)	39	31	18
Bill Bradley(比尔·巴拉迪)	55	34	28

来源:作者根据原始数据档案(2000 American National Election Study)的计算结果。原始数据见 http://www.umich.edu/~nes/.

者变得怨恨和孤立,因为决策者的决策取决于较为冷静的多数支持者。过了几年,少数派变成多数派,才有能力改变政策。

测量偏好的强度,更有创意的方式是用程度来测量意见。有些调查让人们用“情感温度计”上的刻度来评估候选人或政党:0 代表否定,50 代表态度中立,100 代表肯定。这有助于扩大回答的范围,更精确地显示意见的强烈程度,但它仍可能掩盖大量信息。有些人尽管确实没什么意见,但仍可能对问题做出回答。⑥

考虑一下民众对竞选 2000 年总统的候选人有关的“情感温度计”调查问题所做的回答。如表 2.1 所示,对候选人持中立态度的人的比例,似乎与确实听说过该候选人的比例有关。相对不知名的候选人帕特·布坎南和雷普·纳德,得到“50”的评分远多于广为人知的候选人布什和戈尔。这就使一些人怀疑,这个问题能否有效地测量人们对不太熟悉的候选人的中立“情感”,或者能否有效地测量“无态度”或其他态度。

变量的可靠性和有效性

相较于含混语句中的模糊用词来说,恰当构想和实施的变量量化测量,能更有效地详细表明差异和程度。不管我们决定怎么测量变量,都希望找到一种计算方法,即便其他的研究者使用,也能产生可靠的结果。

例如,我们决定问一些随机抽样出的回答者,他们是“喜欢”还是“讨厌”总统(因此是二择一的强迫选择),以此测量“总统的民意支持”。我们可以预期,其他的调查者在下一天使用相同的测量方法,问数量相当的随机样本,将

⑥ 对“无态度”问题的讨论,见 Herbert Asher, *Polling and the Public: What Every Citizen Should Know*, 4th ed. (Washington, DC: Congressional Quarterly Press, 1998), chapter 2。

产生相似的结果。

反之,我们可以问一辆公共汽车上遇到的前4个人,他们对总统"有何看法",然后通过对他们回答的个人印象来评估总统的民意支持。而其他人来日在同一辆公共汽车上采用相同的测量方法,可能产生完全不同的结果。答案可能是含糊的,而影响到答案解释的价值观可能也有差异。

如果不同的人对变量的测量产生相同的结果,那么这种测量通常就被说成是可靠的。强迫选择式的问题很可能产生一致的结果,因为使用相当样本量的每位调查者必须计算"喜欢"和"不喜欢"的人数,来测量民意支持。然而,与公共汽车上的人进行开放式的讨论,需要调查者解释可能(或可能不)反映民意支持的各种评论。这类回答是有意义的,甚至就某些方面而言,比强迫回答更有意义。但是它们不大可能就总统声望这个问题提供一个可靠的答案。

理论上对变量测量方法的关注可以做到很细腻。我们使用的每种测量方法,都被假定很好地反映了一个量化的变量背后蕴含的抽象概念的真实情况。如果一种测量方法"做到该做的",[7]就是有效的。一种量化的测量方法越能反映研究关注的底层概念的定义,它就越有效。社会科学的困难之一是,从来就没有明确的方法用于直接评估有效性。举例来说,智商测试是许多试图量化智力的调查者使用的或许可靠的测量方法。但是,这种测试对于测量智力这么一个丰富的、变化多端的、有影响力的概念,能达到多大程度的精确性,永远是一个有争议的问题。智商测试可能可靠,但是否绝对有效?

构想不当的测量是危险的,正是因为它能带来太大的影响。一个不幸的、令人反感的例子是,"尸体计数"在美国当做关键手段,用于"推进"它在越战上的努力。战争的新闻广播报道通常是军事数字,说每天有多少"敌人"被打死。言下之意是,打死的敌人越多,就能越快赢得胜利。这种量化的测量有两个错误。

第一,它并没有测量某些决策者宣称它已测量的东西:在实现战争的总体目标中胜利或失败的数量。由于战争不只是军事冲突,它至少还是一场政治和心理较量,尸体计数在很大程度上不能作为胜利的一项指标。它们或许能让军方知悉一些敌情,但这种方式会在越南人民和美国人民那里产生不利的政治和心理影响。越南人民开始注意到,无论这场战争为何,被美国打死的主要是与自己同种同国的人。因此,许多越南人开始畏惧美国人,而不是把他们当做盟友来欢迎。与此同时,美国人民也开始把自己看成是给一个贫穷国家带来灾难的行尸走肉的战士。

---

⑦ Edward G. Carmines & Richard Zeller, *Reliability and Validity Assessment.* Sage University Paper Series on Quantitative Applications in the Social Sciences, no. 17 (Beverly Hills, CA.: Sage Publications, 1979).

测量中的第二个错误是它的执行。据说野战部队要计算并报告敌人的死亡数目。然而出现了几个因素：认定敌人上的混淆（有时是故意的），在同一个地点有两个以上的人计数所造成的错误，指挥系统对高尸体计数的要求带来的压力。因此，尽管尸体计数不断上涨，并导致对战争胜利的预测，但实际情况却恶化了。[8]

有一点非常重要：草率的或不当的测量，一般比不测量还糟糕。对测量结果的解释需要对测量本身有所理解。第5章将探讨测量实施和测量解释过程中的实践步骤。

假 说

尽管前面的许多讨论，看起来像是对零零星星的科学思维做系列回顾，但对假说的讨论可以把这些题材组合到一块。

假说是一种特殊建造的句子。假说的目的是组织研究。如果假说得到了精心设计，科学方法的所有步骤也就随之而来，如项目大纲、参考书目、所需资源以及适用于研究的测量方法的说明。假说提供了整个的结构。

假说提出的是两个或更多变量间的一种关系。例如：政治参与随教育增多而增加。这个简单的断言也可看做是一条假说。它有一个主项（变量：政治参与），一个连接动词（一种关系：增加），以及一个宾语（变量：教育）。

我们再看两个例子：

异化随贫穷加剧而增加。

工会成员比非工会成员更有可能投民主党候选人的票。

或者，不那么明显的例子（你可以为练习而从中找出变量及关系）：

距离产生美。

每天一苹果，不用上诊所。

早睡早起，健康、富裕和聪明伴随你。

至关重要的是要认识到，假说其实是一种假定，正如牛津英语词典指出的："作为深入研究的起点，并被证实或证伪……"假说靠近研究的起点而非终点，尽管优秀的研究可能会提示富有成效的新研究途径和新假说。

到目前为止，我们关于假说的大部分例子都是相当简单的。从简到繁，我们来引用一次教导科学思维的经验。有名学生来见一位笔者，他带着下述研

---

⑧ 在1991年的波斯湾战争期间，盟军军事指挥部为避免这些问题，没有公布对伊拉克军队估算的伤亡数。

究提案：

> 殖民者与殖民地居民之间的不稳定关系（形象地说，类似于连体双胞胎的关系），以及他们分别与殖民状况的互动，导致了脆弱的双螺旋型心理病态的发生，它对心理颇有影响，其结果导致心理失调，即各种神经官能症，它们随即确定了殖民地的政治细节的本质。

这仅仅只是提案的开始！在所有这些混乱的语言中，存在很多变量和关系。不过，梳理后可以得到两条假说：

> 殖民主义关系到殖民者和殖民地居民的神经官能症行为。
> 这种神经官能症行为影响到殖民主义的政治结构。

这两条假说尽管很大，但尚可处理。殖民主义这个概念描述了一种定型的政治格局。“关系到”这种关系不像“导致”那么口气强硬——这只是因为学生能获得的研究资源有限，因此要谨慎点。神经官能症行为也是一个棘手的概念，但出自于心理分析理论文献，有些行为确实还能贴上神经官能症的标签。这样一来，真正的工作就是表明，在殖民条件下发生的各种神经官能症的、自毁的行为，关系到殖民政治的压迫与独裁模式。

如果这位学生通过证据搜集、测量和评估，完成了这些假说所涵盖的工作，那么他就可以申请博士学位了。不过大家都知道他谈的是皮毛，他的论文（居然胆敢命题为《殖民主义：神经官能症游戏》）最多也只能说是达到大学生的要求。

这个例子表明了一点，在形成假说之前，往往存在一个先行步骤，称为问题重构。在上例中，我们是从对殖民主义和神经官能症的一般性关注着手。那名学生在阐述这一关注时使之复杂化。而我们通过厘清变量和关系，把它简化，使之能得到处理，至少能用一般的方式来处理。采用一种可行的重构后，要定义变量的表征方式，就变得比较容易了。

社会科学的艺术之一，就是有技巧的问题重构。除了一些分析常识外，重构还需要一种能力，能看出环境中的变量和变量间的可能关系。第一步最好是把问题分解成作为其构成部分的变量和关系。写下与问题有关的系列假说，这样就能从中挑选出可以回答下述两个疑问的假说：对于解决整个问题而言，哪些假说至关重要？就你所掌握的资源范围来看，哪些假说有信息支持？有时候，这些问题会迫使我们做出一些不愉快的选择，但可以防止研究工作还没有什么实质性的结论就结束了。上面关于殖民主义和神经官能症的例子，就清楚地表明了这一点。

在启动研究任务前，恰当地提出一条假说，其重要性再怎么高估也不为过。

下述原则会对你有所帮助：

1. 变量必须得到明确说明，并能用你掌握的技术加以测量。
2. 变量间的关系必须得到精确陈述，而且必须是可测量的。
3. 假说应当是可检验的，这样，表明关系的那些证据才能被观察、证明或证伪。

如果不遵循这些规则，假说可能是行不通的、荒唐的，或是在现有资源下无法研究的。简而言之，对测量的精确定义和详尽说明是关键。谨慎构造假说的努力或许没什么乐趣可言，但在这个过程中出现的问题迟早要予以回答。

因此，假说提供了整个研究工作的结构，无论该工作是访谈调查，还是分析以前收集的数据，抑或是文献研究，或三者兼具。假说将把你直接引向相关的资料，避免你做无用功。利用图书馆的搜索引擎、互联网、图书索引、期刊网上导读服务，以及只读光盘/电脑数据库检索，你可以对你选择的变量加以研究。变量之间所设定的关系，指示了你需要使用的测量工具和评价标准。假说的检验结果就是你的真正结论。

一旦通过构造和检验假说，建立了变量间的关系，这些关系就能表达为概推。以得到检验的关系为基础的概推，就是科学的目标。概推就是得到检验证实的假说。随着某一研究领域中概推的积累，它们就构成了理论的原始材料。不过谈这还为时过早。现在，我们需要看看科学方法如何把研究程序变成一种逻辑秩序。

科学方法

科学方法这种技术其实很平常。模型研究有一些步骤，其中包括：

1. 鉴别需要研究的变量。
2. 提出有关变量关系或变量与环境关系的假说。
3. 现实检验，以测量变量中的变化，从而观察被假设的关系是否受证据支持。
4. 评估：把变量间被测量的关系与初始假说进行比较；提出有关发现的概推（关系）。
5. 就有关发现的理论意义、检验中可能歪曲结果的有关因素，以及研究可能引出的其他假说提出建议。

尽管我们在这里已经勾勒出科学方法的骨架，但实际研究程序并非总是直接起始于假说的构造。在陈述假说之前，社会科学家通常会检查某一领域中所收集到的资料，以便观察变量之间是否存在关联。各种统计程序所揭示的关系通常能提供线索，指出一些很有探索价值的假说。有时候，只是翻翻一些资料，就能促发一个精彩的思想、一次偶然的洞见或一种全新的观念。过去

数十年来已积累大量的资料，所以研究者在做每项研究时，通常用不着白手起家。现有资料的分析在鉴别那些需要用于检验关键关系的新资料方面极为有用。

这也仅仅是科学方法的轮廓。熟练的分析者还会引入其他的元素，比如：测量结果时对替换法（alternative forms）的使用、对变量描述的详细概念分析、自己的研究与其他研究之间的关系、对测量工具有效性的评估、实验组和控制组的使用，此外还同等重要的是，对已检验之外的东西做出谨慎地猜测。不过，方法论上的这些装饰，更紧密地关涉到实施方法时使用的工具，而非方法本身。

要点是，科学方法是利用可观察的证据，以一种训练有素的方式来检验思维，并且在该过程的每一步都做到明晰。

考虑下述两类研究的差别：（1）一项经验性的科学研究：研究者陈述其价值，构造假说，设计检验程序，仔细选择和讨论测量方法，产生一种特定的结果，并把该结果与假说关联起来；（2）一项非科学的研究：研究者表达其价值，提出一般性的论题，考察相关的例子，并陈述结论。

注意：两项研究中都存在思维和调查之间的张力。但一个重要的区别是，第一项研究可以重复，能对结论的有效性进行检验，而第二项不行。社会科学家用复制这个词表达一种能力：研究能被重复，以检验其有效性。复制构成了对优秀研究的强检验，因为它能揭露出研究程序中不知不觉渗进的一些错误和主要研究中包含的评论性判断。

非科学研究的第二个困难在于，它在把两项研究关联起来时存在着问题。你曾在一次讨论中因别人要求你“定义你的用语”而感到苦恼吗？你曾陷入一场以“你怎么知道它是正确的”为结尾的争论吗？

要弄清楚所发生的现象需要一些信息，好的科学研究就能提供所有这类信息。例如，如果采用标准的变量定义，而且我们就选举人对候选人的评价做了一项研究，那么该研究就可以被吸收进其他的研究：如选举人如何看待有关问题、政党，等等。随着科学家努力构筑积累性的知识体系，使用不同测量方法对同一变量所做的不同研究就能得到比较，看看测量技术是否产生一些不同的结果。同样，要点在于：科学调节、阐明思维与调查之间的关系，并能让其他人确切地知道该项研究做了什么。

理论的多种角色

科学的权威源于它坚实的基础：关于“现实”的可观察证据。有时候我们就是把科学描绘成现实检验。由于每个人都认为自己清楚什么是现实，所以

科学有根本的号召力。然而,科学中实在论的必然搭档是一种完全想象性的现象:理论。如果没有理论所扮演的多重角色,就不会有科学(另外,还有人会争论说,同样也不会有可以理解的"现实")。

正如语言产生于满足人类需要的经验,理论也源自于人类所面临的任务。最艰巨的任务是说明究竟是怎么回事。关于理论是什么,不是什么,已经有大量著作。对于我们来说,理论就是一套互相关联的命题,为事件的发生方式提供说明。

构成理论的命题与假说一样,具有相同的形式:它们都是由概念以及概念间的联接或关系构成。一旦假说得到检验,出现了新关系,理论就构筑起来了。

最平常的日常事务是生产理论的沃土。从自动贩卖机付款到基尔博特动画片的深层含义,都有相关的理论。最宏大的理论是宗教和哲学,它们关注各种各样的大问题:宇宙起源、物种历史、生活目的、导向美德的行为规范,或许还有幸福。对于忠实的信徒来说,这些理论的"真"取决于对超自然现象的信念。这类理论在提出时,就好像它们深植于我们的生存这个更广阔的宇宙中,静待我们的理解。

相反,社会科学对理论持一种不同的看法。社会科学家最常见的态度是实用主义的:只有在说明观察时理论具备当下的或潜在的用途,才说得上好。任何科学的要旨都在于提出一套理论来说明其观察范围内的事件。

认为理论就是变动不居的日常经验背后静止不变的东西,这是一种很有吸引力的想法,但是是误导人的。事实上,理论是人们在追求理解时偶然的天才性创造。人类根据需要创造理论,而这些理论又能反作用于那些需要,对之产生有利或不利的影响。一个理论可能包含一套完整的范畴和概推——但可能还是没什么作用。例如,假如有个人把世界分为高的物体和矮的物体,并把它们之间的所有关系都做了刻画,一个理论就诞生了,但它的效用是可疑的——不是错误,而是没用。

社会科学的理论通常源于一些关于人类行为的基本假定。理性行动者理论认为,个体、组织和国家都以最大化他们的物质利益为动机。基于该理论,我们可以猜测,投票人选举候选人是为了增进他们自己的经济利益。然而,心理学理论断定,投票行为是由人们对政党的长期感情决定的。该理论认为,投票人被社会化,由于家庭的影响而忠于某一个特定政党。基于该理论,我们可以猜测,投票人的行为与其父母的行为是一致的,或者说,他们会年复一年地选择同一个党派。理性行动者理论、心理学理论、马克思主义理论和其他形式的社会理论都对战争的起源做过说明。

我们讨论的是理论是什么,不是什么。下个问题是:理论可以用来做什么?答案是:许多事情。我们可以列举理论在社会科学思维中的四种特殊作用:

1. 理论为资料解释提供了模型。
2. 理论把研究彼此关联起来。
3. 理论为概念和变量配备了框架,从而使之获得实质意义。
4. 理论使我们能向自己和他人解释我们的发现的深远意义。

我们且通过考察投票人参与这个问题来阐述理论的这四种作用。投票人参与的比例是民主的一项重要指标。我们有理由预期,不同类型的选举规则会影响那些认为值得花时间和精力去投票的人数。我们将表明,理论会如何影响到我们看待对立选举规则下政治参与问题的方式。从资料中观察到的模型、研究之间被建立的关联、有关发现的实质含义,以及它们的深远意义,都是由研究者使用的理论塑造的。

在这个例子中,我们将集中关注,选举规则是怎样把人们的选票变为政党的议席的。在美国,几乎所有的选举都支持"单一选区制"。这些规则为每个选区中获得多数选票的候选人提供一个议席。假如这些规则允许更多的政党获得议席(比如在比例代表制中),那么就有更多的人可能参加投票。[9] 在比例代表制中,每个政党基于其得票比例而在总议席中赢得相应的议席。比如说,如果在一次选举中,分配了 10 个议席,那么其中大部分席位可能会落入大型政党的候选人手中,但如果小型政党能赢得 10% ~20% 的选票,就可以推举 1 ~2 个席位的候选人。

有人提出理论说,在单一选区制下,赢者通吃的规则可能会减少投票者的参与。在美国,这样的规则意味着,几乎所有的议席都会被民主党和共和党的候选人占据。其中的假说就是:不偏向主要政党候选人的公民可能不愿意投票。[10] 由于很大一部分选民视自己为"中立派",所以这就成了评估美国民主实效的重要因素。

考虑一下表 2.2 中的数据。该表显示了三类选举规则下不同国家进行选举时投票人数的平均水平。赢者通吃规则只把议席赋予单一选区中率先胜出的候选人。比例代表制一般允许投票者选择一个政党的候选人名单,然后大致按照每一政党的得票情况分配多数席位。有些国家采用赢者通吃规则选出部分席位,而其余的采用比例代表制。这些国家属于混合体制。

---

⑨ 例如,参见 G. Bingham Powell, "American Voter Turnout in Comparative Perspective," *American Political Science Review* 80, no. 1(1986), pp. 17-43。

⑩ 有项研究试图保持文化差异不变,比较在美国采用赢者通吃规则的局部选举时的投票人数与采用"半按比例投票制"时的投票人数。该研究发现,"半按比例代表制"下的投票人数增长了 5 个百分点。见 Shaun Bowler, David Brockington & Todd Donovan, "Election Systems and Voter Turnout: Experiments in the United States," *Journal of Politics*, 63, no. 3, (2001), pp. 902-915。

表2.2 20世纪90年代国家选举投票情况

	议会的选举规则类型		
	赢者通吃[a]	比例代表制[b]	混合型[c]
平均投票人数	66.5%	74.3%	77.3%
标准差	9.4	11.3	3.8
国家数量	4	3	12

[a]加拿大、英国、法国和美国

[b]德国、日本、新西兰

[c]丹麦、芬兰、希腊、冰岛、爱尔兰、以色列、荷兰、挪威、葡萄牙、西班牙、瑞典、瑞士

注:平均参与投票数来自19个发达的工业民主国家的大选,其中投票大多出于自愿。

来源:作者对原始数据的统计,数据来源于 Russell J. Dalton, *Citizen Politics: Public Opinion and Political Parties in Advanced Industrial Democracies* (Chatham N. J.: Chatham House, 2002), p.36.

这些数据传达了怎样的信息呢?先看最上面一行的数据,我们可以发现,选举规则会影响到国家选举的参与情况。采用比例代表制的国家投票率达到74%,相比之下,实行赢者通吃规则的国家投票率只有66%。这些数据是不同国家的投票情况的平均值。虽然不能直接证明,但它们也暗示,比例代表制使更多的公民参与选举。然而,一些竞争性的解释以及干扰性的变量,如文化差异,也可以说明这些模型。⑪

理论在哪?什么样的理论适合这一数据模型?一种理论是说,当人们认为其行为会产生实质性的结果时,才更有可能去行动——在这个例子中是投票行为。换句话说,如果他们认为他们所支持的政党会获胜,或者认为在势均力敌的选举中他们的投票很重要,才更有可能去投票。因此,比例代表制可能吸引小型政党的追随者去投票,因为他们的投票更有可能影响到某个代表的选举。由于比例代表制使更多的政党可能获得立法席位,所以表2.2中的数据似乎更支持下述理论:比例代表制能调动更多的公民,因为他们的投票更有可能产生实质性的结果。

另一种理论认为:人们投票是出于"公民义务"感。根据该理论,他们投票时就不会考虑他们支持的政党赢得议席的几率。⑫ 由于没有数据表明公众的义务感在这些国家之间是如何变化的,因此,尽管我们的数据与一个理论相

⑪ Douglas Amy, *Real Choices, New Voices.* (Cambridge, England: Cambridge University Press, 1993), pp. 140-153; Arend Lijphart, "Unequal Participation: Democracy's Unresolved Dilemma," *American Political Science Review* 91, no. 1 (1997), pp. 1-14.

⑫ 关于这些相争理论的讨论,见 Donald Green & Ian Shapiro, *Pathologies of Rational Choice Theory* (New Haven, Conn,: Yale University Press, 1994). Cf. Kristen Monroe, *The Heart of Altruism: Perceptions of a Common Humanity* (Princeton, N. J.: Princeton University Press, 1996)。

吻合，但该事实并不意味着我们就能否认另一个和它相竞争的理论。

留意不同的理论，能使社会科学家把其研究同先行的研究关联起来。它还提供一种方法，制造额外的检验，从而使我们可能拒斥那些为数据模型提供其他说明的竞争性理论。例如，如果我们发现"公民义务"在表2.2的数据所反映的地方没什么差别，那么我们就有更强的理由认为，比例代表制比赢者通吃更能调动人们去投票。

与投票所付出的代价相比，比例代表制可能带来更多实质的好处（代表的更大机会）。情况有可能是，在美国，不投票的人感到在政治上被边缘化了，因为选举制度阻止了他们支持的候选人获胜。利用推理中的这些关联，社会科学家就能积累有关不同理论建构之关系的知识。至此，我们联系表2.2中的实例看出了理论的两种作用：理论提供的模型，以及理论把一项研究同另一项研究关联起来的方式。

现在，理论的第三种作用就显而易见了。在这里，我们需要评估观察的实质意义。⑬ 也就是说，我们需要追问，观察是否有一些精彩的、重要的寓意。对于检验"理性行动"理论在说明政治行为方面的效用，这个结果可能很重要。就这个例子来说，我们可以推断，在一个采用赢者通吃的国家，改用比例代表制能提高10个百分点的投票率。⑭

看来，选举中给公民越多的选择，参与的人可能就越多。这就引起了一系列有趣的、实质性的问题：这些投票者的参与将如何改变政治体制？什么样的新政党可能获胜？类似于国会这样的机构如何应对几个政党？

对于把政治机构与人类行为联系起来的理论来说，这些发现的深远意义超越了那些具体的实质性问题。代表民主制的参与，不仅仅是有形式上的投票权，人们同样对该过程的结果和有关制度产生的限制非常敏感。当然也有其他的因素卷入，不过，那些产生更多社会团体代表的选举制度，似乎也能鼓励更多的人参与投票。比例代表制能产生出更广泛群众的代表，一般而言，这种制度同样可能更广泛地影响到公民对政治和政府的态度。⑮

---

⑬ 不应该混同于统计意义（见第5章），后者告诉我们的是，赢者通吃和比例代表制之间的差别是不是偶然的。实质意义与理论相关，而非与统计概率相关。

⑭ 一些解释额外变量的比较研究认为，比例代表制对于投票率的独立作用效果在3%～7%浮动。见 Andre Blais & Agnieszka Dobrzynska，"Turnout in Electoral Democracies，" *European Journal of Political Research* 18（1998），pp. 167-181；Andres Lander & Henry Milner，"Do Voters Turn Out More under Proportional Than Majoritarian System?" *Electoral Studies*，18（1999），pp. 235-250。

⑮ C. Anderson& C. Guilloty，"Political Institutions and Satisfaction with Democracy：A Cross-National Analysis of Consensus and Majoritarian System，" *American Political Science Review* 91，no 1（1997），pp. 68-81.

在讨论理论的过程中，我们用实例表明了它在社会科学研究中的作用。大多数研究者都想证明其理论是“对的”。然而，一位颇有影响的社会科学分析家卡尔·波普尔(Karl Popper)表明，科学的最佳用途通常是反驳理论，而非“证明”它们：

> 几乎每种理论都可以说是符合许多事实的。这也是为什么我们能说，只要不能找到反驳性的事实，而非不能找到支持性的事实，理论就得到了确证的理由之一。[16]

换句话说，数据给人留下的印象，更深刻的是作为反驳理论的证据，而不是作为支持理论的证据。

在对理论的讨论中，我们没有把握到的是人们在思考其观察对象时的微妙性和创造性。理论启发观察。然而，就像照在物体上的一束灯光一样，每种理论都会留下自己的影子，挑战我们的想象力。

一方面，我们只能说，没有理论，社会科学就会沦为没有条理且毫无意义的一堆观察、数据和统计数字；另一方面，并不是所有的社会科学都限定于严格的、明确的理论表达。不过，绝对清楚的是，复杂的社会问题需要我们尽可能做出材料充分翔实的研究。用理论的形式组织和评估知识，几乎与作为第一步的知识搜集一样重要。历史上到处都是构造糟糕的社会理论残骸，有时候产生悲剧性的结果，尽管有些最伟大的文明进步归功于理论想象的能量。

现在我们掌握了科学思维的基本工具。但是工具本身不能完成工作。我们还需要一套方案，或者，就如下一章描述的，一种策略，把工具派上用场，来获得某种知识。

引入的概念

概念(comcept)
变量(variable)
不连续量化(discrete quantification)
连续量化(continuous quantification)
测量(measurement)
可靠性(reliability)
有效性(validity)
假说(hypothesis)
问题重构(problem reformulation)
概推(generalization)
科学方法(scientific method)
复制(replication)
理论(theory)

[16] 见 *Popper Selections*, ed. David Miller (Princeton, N. J.: Princeton University Press, 1985), p. 437。

讨 论

1. 考虑概念“失业”。

- 如何定义它，使之能作为一个变量被测量。
- 你能想到多少个“失业”的定义？
- 这些定义有何不同？

2. 从可靠性和有效性的角度评价一下测量失业的不同方法。

- 就可靠性来说，如果其他研究者采用你的测量方法（变量），会产生相似的结果吗？
- 从有效性来讲，这些测量方法能很好地反映“失业”这个概念吗？

3. 考虑一下 1996 年大选中关于宗教和投票行为的数据。制作一个表格，反映这两个变量间的关系。你能就宗教与投票行为的关联方式（或为何有关联）提出假说吗？

4. 社会科学家处理的更为复杂的问题之一是：人们为什么要反抗他们的政府？考虑三次大型的革命例子（17 世纪的英国，18 世纪的法国和 20 世纪前叶的俄罗斯）。以这些例子为基础，你能提出革命发生原因的假说吗？在提出假说时，考虑一下：

- 革命的发生与哪些变量有关？
- 你如何定义如革命这样的概念？
- 其他的研究者能否把你的变量可靠地应用到其他发生或没发生过革命的国家？
- 你假说中的变量是如何被关联在一起的？
- 你会怎样检验你的假说？
- 如果别人使用你建议的测量和检验方法，能得出和你一样的结论吗？

第3章 策略

Strategies

"事实就像袜子，空的时候不能立起来。要让事实立起来，你得往其中加进使之得以存在的理性和感受。"

——路易吉·皮兰德娄(Luigi Pirandello)

敏锐的读者会注意到，通常被认为是科学方法所必需的两个词，在本书里却很少出现。它们就是事实和真理。这两个词的共同之处是：它们都带有一种绝对主义的意味，误导那些着手用科学方法来获取知识的人。从字根来看，事实意味着"一件已完成的事情"。已经做完的事情是无可争议的，但问题在于，"已完成的事情"并不是由某个中立的全能观察者而是由人感知到的。

人只有有限的感知能力、本能和利益结构，这都影响到他们观察世界的方式。科学是一个过程，尽可能使这些感知变得明晰，并能不受限制地接受检验。但是我们必须始终这么看待科学程序的结果：它是一种努力，旨在控制一个过程——即便并非不可能，但也被我们的人性弄得难于完全控制的一个过程。

出于工作目的，社会科学家通常把事实看做是"根据理论目的而对现实所做的一种特殊安排"。[①] 任何被定义为事实的事情，都是与观察者在研究现象时所持的特定利益联系在一起的。跨过这一步是不会有什么作用的，因为哲学丛林就在前方隐现，里面蕴藏深奥的问题：倒下的一棵树没被观察到，既然我们无法知道它确实倒下了，那么它究竟倒下没？

对哲学家来说，真理这个词就像块牛肉，大受他们的欢迎。科学更倾向于在不那么高贵的领域里工作，这个领域由一些可证伪的陈述构成，它们能得到其他人的检验。对好的科学命题或概推的陈述，要能让随后的观察提供支持性的证据，或是提供质疑该命题精确性的证据。如果把证实程度作为科学的

---

① David Easton, *The Political System* (New York: Knopf, 1953), p. 53.

一个持久考虑，就可以避免大量草率的结论。

对事实和真理所持的这种不明朗的态度，有人对它的反应可能是追问："那么，我们究竟要相信什么？"如果你想要绝对可信的东西，那必定要到科学以外的领域去找。科学是一种工作程序，利用对经验的提炼来回答问题。科学家可能会提出能量惊人的理论，但这样的理论能否满足我们人类的信仰需要，则是一种私人事务，无关于科学探索的意义。"相信科学"就意味着接受一些判断，它们建基在能被复制的观察之上，而不是其他类型的证据或心理过程之上。

现在你已经熟悉了科学的基本元素，比如变量、测量和假说。本章我们将集中讨论的是，如何给我们关于世界的观念配备一种形式，使之能接受现实的检验。接着，我们把现实检验过程分解成几个部分。最后，我们将看看，为了理解研究结果，需要采用哪些评估步骤。

以下各项的设计，旨在为科学分析提供逐步的指导。不过必须意识到，我们只是试图把握住科学程序中最重要的方面，而不是老到的研究者想引荐的那些繁枝末节或是错综复杂之处。下一章以"提炼"为题，会对每个要素补充一些看法，以增强你在研究策略上的能量。

请记住，我们在这里所做的一切，都是调节人类思维的自然状态：思维与现实检验之间存在的一种张力。所以这一章分为三个部分：仔细思考问题、现实检验和理解结果。

仔细思考问题

做研究时的最大挑战来自于起始阶段。一旦你遭遇挑战，其他步骤随之而来。难题就在于限制主题，或更明确地说，就是选择一条通向主题的途径，它要能最有效地使你到达你想要理解的东西。大多数学生都有过写一篇冗长、散漫、没什么主题的论文的经历。到最后需要下结论时，有时候却尴尬地发现，从已经提供的证据来看，没什么坚实的结论可下。出现这种丧气的结果，其原因通常能在起始阶段找到。

焦　点

由于我们大部分人没受过这样的训练，即用假说的形式阐述我们的观念并检验它们，因此，我们最好从现在开始，按照事物出现在心灵中的方式来记下它们：连串的主意、想法和观念。追问自己："我为什么对这感兴趣？我究竟想要什么？"看看有什么发生。你可以从一个广泛的主题开始。

这个国家有个大麻烦。大多数人现时觉得政治不重要——他们不想

与政治有任何瓜葛。政治就像个笑话。

多大的一些问题啊！不过这里还有一个主题：现代民主是否奏效。

在此阶段，用一个或两个段落来记述这些思想是个不错的主意。写在纸上！广泛地阅读资料也是个好办法，它有助于圈定调查的领域。但读得太多却可能是个糟糕的做法。在试着进行实际研究之前，要对问题的主体框架做通盘的考虑。

假设你以类似下述两段的内容结束：

在美国有些事已经变了。以前，似乎更多的人认为他们在一些事情上有重要的影响。也许他们认为政府可以把社会改造得更好，或者他们可能有更多的闲暇时间。

如今，与我交谈过的大多数人都不关心政治，也不相信政府。谁有时间理会政治？另外，除非你有钱捐助，否则他们不会听你的。

这些段落实际上包含一些概念和变量，一张关系网和一整套假说。至少有迹象显示，更为集中的研究是有可能的。

这个时候，我们可以开展两层研究：描述性的和关系性的。描述性研究搜集与一种状况有关的信息。我们可以描述一种制度、事件或行为，或者它们的组合。好的描述是科学的开始。达·芬奇(Leonardo da Vinci)对人体结构的精彩注解和描绘，促进了几代医学专家对身体的理解。有些专门化的描述性研究所分析的是一个单一变量的有关信息，比如说，家庭破裂。它是由什么构成的？它发展到什么地步？情况是怎么发生的？什么时候发生得最频繁？对于更高级的分析形式来说，这些研究都是有用的资源。

关系分析考察事物之间的联系。其基本形式是探索两个变量之间的关系，比如说，信任和政治参与之间的关系，或是年龄和参与意识之间的关系。系列的关系研究能构成因果分析的基础；因果分析是一类特殊的关系研究，它把变量之间最有力的关系独立出来。

前面几段对相关主题的初步考虑，似乎蕴含着一整套的关系。如果你急于刨根问底，那么下一步可能是做一种关系分析，分析参与和信任这个一般性问题的某个方面。

假说形成

由于主题窄了些，形成假说就变得更容易了。问题有两层：有哪些关键变量？它们之间有何关系？在我们的例子中，引起兴趣的一个部分包含着两个变量：政治参与和信任。前面的那几段话暗示了两者之间存在一种关联。这个关联究竟是什么？有什么词语可以表达那种关系？如果我们不考虑因果分析，那么暗示的关系就很简单了：人们更愿意参与政治，如果他们信任政府，或者，如果他们认为这样做会有实质性的效果。

即便使用了这些词语，但我们终究还是把情况简化为两个变量和一种关系：参与关系到信任。当然，大多数研究包含几条假说，它们可能被贯穿在一起，成为一个大型主题的构成元素。但是出于说明的需要，我们只讨论要求不那么高的东西。

操作概念

操作概念，意味着给它配备一种形式，使我们能用某种方式对变化做出测量。第2章中，我们讨论过转概念为可用的变量，这个过程称为操作化。转概念为某种允许我们对变化做出观察的东西，是一个棘手的过程。要做得妥当，需要满足两个条件：(1)可操作的变量要尽可能符合原始概念的意义(有效性)；(2)对变化的测量能被他人复制(可靠性)。

人们是否参与政治活动，对此问题我们该如何使之可操作化？试试看问人们一些事关政治参与的简单问题：

- 你是否参加过政治会议、集会或演说？
- 你是否为某个政党或候选人工作过？
- 你是否就某件公众事务写信给报刊？
- 你是否在上次选举中投过票，或是做别的事去了，从而避开了投票？

一旦给出答案，要做的就是在各个问题的答案中找出模型，或是评估一下，如何把这套问题的回答合到一起。我们可以审视一下这些回答，看看其中会出现什么样的主题或趋势。无论我们怎么去研究，答案总会提供一些线索，指示出有多少人在政治上比较积极。

对政治参与问题的回答中蕴含的趋势，同样有助于评估我们的研究问题。如果这些问题被问了很长一段时间，我们就能看出现今人们参与政治是变多还是变少了。对政府的信任这个变量通常是由下述问题的答案来测量的：

> 你认为你对华盛顿政府行事正当的信任能维持多久——一直、大多数时间，还是偶尔？

回头温习下前面的几段，我们看到，主要的假说牵涉到，低程度的信任如何可以导致人们现今更不愿意参与政治。规范的民主理论认为，健康的政治不仅要求某种最低程度的信任，也要求某种最低程度的公众参与政治进程。

在这点上，我们将看看一位哈佛的政治学家罗伯特·普特南(Robert Putnam)在研究中实际使用的研究策略。本书的附录A是普特南的一篇论文

的节录版,其文章题为《调来调去:美国社会资本的奇怪消失》。[②]

普特南灌输给我们的一个观点是:低程度的政治参与部分是因为社会团体中成员身份的减少。普特南在提出其理论和假说时,参考了早期的美国社会观察家亚力克·德·托克维尔(Alexis de Tocqueville)和当代社会学家詹姆斯·科尔曼(James Coleman)的著作。[③] 他论证说,由于缺乏信任,比起20世纪50年代而言,人们如今更少参与政治了。

按照一般的预期,在自愿组成的社会团体中,成员身份会对人们有所改造,使他们更多地参与政治,创建民主所必需的"公民能力"。[④] 该理论还说,早几代的美国人更积极地参加一些团体,它们有助于建立社会信任和社会关系。如今,人们可能不那么愿意认识他们的邻居,不那么愿意与朋友打交道。[⑤] 这样的关系使他们能看出"为互利而合作"的作用。简单说,这个论证就是认为:社会活动的减少造成了信任和政治活动的减少。如果我们不参加一些像家长教师联合会(PTA)或是保龄球联盟这样的团体,就不能学会彼此信任。我们因此也就没了解到,如政治这样的集体性活动会有什么实质效果。普特南和其他人把我们从团体活动中获得的社会关系、规范和社会信任统称为社会资本。[⑥] 科尔曼把社会资本定义为"提供帮助"的个体之间的社会关系。[⑦]

阅读附录A中普特南的文章;我们将探讨他是如何做研究的,这有助于提升你的理解。这篇文章是为一项更大的研究计划而精心制作的一个概要。作者开始讨论了工作的理论背景、提出和检验假说的步骤,以及其结果的深远意义。这是研究应该遵循的模式——即便是一篇简要的研究论文。

在操作变量政治参与的过程中,普特南考察了对几个问题的回答,其中包括上面列举的一些问题。他同样感兴趣的是,政治参与的减少如何关系到信

---

② Robert D. Putnam, "Tuning In and Tuning Out: The Strange Disappearance of Social Capital in America," *PS: Political Science and Politics*, 28, no. 4, (1995), 664-683. 另外见普特南的另一篇文章,"Bowling Alone: America's Declining Social Capital," *Journal of Democracy*, 6, no. 1 (1995), 65-78.

③ James Coleman, "Social Capital in the Creation of Human Capital," *American Journal of Sociology*, 94S: 95-S120 (1988).

④ 另一位观察家写到:由于团体身份而带来的"人际关系改造","不可能轻易地用笨拙的社会科学工具加以测量"。Jane Mansbridge, "On the Idea that Participation Makes Better Citizens," 见 *Citizen Competence and Democratic Institutions*, eds. Stephen Elkin and Karol Soltan (University Park: Pennsylvania State University Press, 1999), 291.

⑤ Robert Putnam, *Bowling Alone: The Collapse and Revival of American Community* (New York: Simon and Schuster, 2000).

⑥ Putnam, 666.

⑦ James Coleman, *Foundations of Social Theory* (Cambridge, MA: Belknap, 1990), 304.

任以及社会资本。就像社会科学中的许多概念一样，社会资本并没有为操作化提供一种直接的、明显的手段，它不是我们能直接看到或计算的东西。它是一个很大的概念，普特南的理论只是为我们要寻找什么提供指导而已。他认为社会资本是由社团的成员身份带来的。这意味着，对社会资本的间接测量，可以是人们就可能参与的社团有关的一些简单问题所做的回答。

普特南把社会资本当做社团的成员身份来操作。他用一些调查问题的回答作为其工具之一，来测量团体的成员身份。他的理论引导他用下列团体中的成员身份作为社会资本的指标。

表 3.1 14 个国家的自愿团体成员身份

国家	体育团体/俱乐部(%)	教会团体(%)	文艺团体/俱乐部(%)
比利时	21.9	8.9	14.0
丹麦	34.8	20.9	10.3
法国	16.1	6.2	6.3
德国(西德)	27.9	15.9	4.9
英国	23.8	19.4	6.8
希腊	6.5	1.8	5.6
爱尔兰	25.6	17.5	3.8
意大利	10.2	8.8	6.5
荷兰	35.2	26.9	10.7
新西兰	47.7	38.9	n/a
挪威	27.2	10.5	8.2
葡萄牙	11.5	5.6	4.6
西班牙	8.3	5.9	3.3
美国	21.6	33.4	9.8

注：假定团体成员身份是反映社会资本水平的指标之一。每一个单元里的数字表示自称是成员的那些回答者的百分比。

来源：作者对原始数据文件的分析——Europe, 1990 Eurobaromerter Survey 34.0; New Zealand, 1999 New Zealand Election Study; USA, 1994 General Social Survey。

- 教会团体
- 体育团体(足球队、垒球联盟、保龄球联盟)
- 文艺协会(戏剧团体、乐团)
- 工会
- 友爱组织(狮子保护协会、麋鹿保护协会、共济会、国际青年商会)
- 服务团
- 民间组织(童子军、红十字会、家长教师联合会)

表 3.1 表明了人们宣称自己是这些团体的成员的频率。尽管普特南主要关注的是美国的趋势，但这些描述性的数据表明，相较于其他许多民主国家来说，美国人参加社团的比例与其他国家相当或更高。现在，我们用团体成员身

份操作了关键变量社会资本，因此，整个研究的组织工作已经准备就位。

现实检验

整理参考目录

有了假说之后，在开展实际研究之前，最好是多读一些资料。这使你能在其他人的工作的帮助下，检查你的假说形成和变量操作。使用图书馆的数据库或互联网来查找假说中的变量。论文、著作和网址都是能提供信息和背景的有用资源。通常来说，就相关主题发表的一篇期刊论文会有一些脚注和参考目录，这能让你知道与该主题有关的大多数最重要的文献。更老练的研究者会在开始就走这一步——它有助于我们在“仔细思考”这个阶段节约大量的时间。然而，初学者在遇到社会分析问题时，经常显得很“冒失”。

开展研究

很多学生也许没想到开展这里建议的那种研究工作。如果你想做深入研究，这里提供的调查数据可以在互联网上找到。[8] 下面的例子是用作指导，不过它们应该能够表明社会科学研究是如何进行的，帮助你能用最清楚的可能策略来阐述你自己的研究计划。

罗伯特·普特南以及在此领域工作的其他人，通过重新分析过去的调查来检验他们的假说。许多社会科学就是这样开展的。当研究者站在不同的角度来审视“老”数据、提出新假说时，就揭示出了新颖的洞见。在你做研究的时候，与同行协商，检查一下图书馆，看看在着手收集你自己的数据前，是否有你能用于检验假说的数据。

在这种情况下，学者们通常查看综合社会调查（GSS，对美国人的态度、舆论和行为的一项主要的学术性调查）中提供的数据，来考察社团成员身份的趋势与信任及政治参与的趋势是如何关联起来的。这不是普特南使用的唯一数据来源。他还参考其他的调查和团体保留的成员记录。

在其理论的基础上，普特南提前认定，某些类型的团体成员身份在建立信任和促进政治参与上比其他的更重要。宗教团体、工会、家长—教师组织、民间团体和友爱组织是比较突出的——但是他的数据来源为许多其他类型团体中的成员身份提供了测量方法。

---

⑧ 公众舆论和选举行为的国家选举研究导引，见 http://www.umich.edu/-nes/。综合社会调查存档见 http://www.icpsr.umich.deu:8080/GSS/homepage.htm。GSS 网上允许对调查数据做在线分析。罗伯特·普特南的书中用到的数据见 http://www.bowlingalone.com。

普特南在其研究中讨论的调查问题，在近 20 年里每年都会被问到，这样他就能比较团体成员身份、信任和政治参与的趋势。总结对这些问题的回答，可以揭示出某些有趣的模型（见表 3.2），也能允许做一些有意思的观察：例如，一些政治参与形式在减少，但是选举似乎仍很火。政府记录实际上表明了投票者参与的减少，那么对选举问题的回答是否可信？[⑨] 我们同样也看到，对政府的信任也有实质的下降——参与减少与信任降低有关吗？

在任何情况下，表 3.2 中的结果尽管可能对于美国社会来说不太乐观，但似乎确实支持该项研究的一些基本观点——政治参与和信任都在下降。不过，比起简单地说“瞧，我是对的（或错的）！”一项研究的结论还应包括更多内容。

分析结果

我们应该正确地分析结果。在这项研究中，真正的问题是，某些特定社团的成员身份是否更能促使人们参与政治和信任政府。我们有很多方法可以检验这种关系。其中一个方法是考察一段时间内的趋势。图 3.1 根据 GSS 和 NES 调查问题的回答中体现出的趋势，提供了一些初步结论。我们看到，从 1974 年到 1994 年，教会团体的成员身份处于下降趋势（尽管只是稍微下降），回答者中，对政府的信任也从 1974 年的高于 40% 下降到 1994 年的 20%。这些结果似乎支持下述基本假说：社团的成员身份与对政府的信任有关——但它们必须接受更深入的分析。如表 3.2 所示，从 20 世纪 70 年代以来，对政府的信任实际上相当稳定（并且很低）。自 1960 年后，自认的政治参与者似乎也有轻微的下降，但我们能肯定这是因为社团成员身份的减少而引起的吗？

表 3.2　1960—2000 年政治参与和信任的趋势

	1960(%)	1968(%)	1976(%)	1984(%)	1988(%)	1996(%)	2000(%)
出席政治性会议、集会或演讲	8	9	6	9	7	5	5
为政党或候选人工作	6	6	5	6	3	2	3
在上次大选中投票	79	76	72	74	70	73	73
一直/大部分时间信任政府	73*	61	33	44	40	33	44

* 1958。这个问题不是在 1960 年问的。

来源：国家大选研究，不同年份。

⑨ 出席总统选举的实际投票比例从 1992 年到 2000 年徘徊在 50% 左右。NES 的调查很可能过多地抽样实际的投票者。与此同时，样本中一些实际上没有参与投票的人可能告诉采访者他们投过票。我们将在第 5 章讨论这些问题。

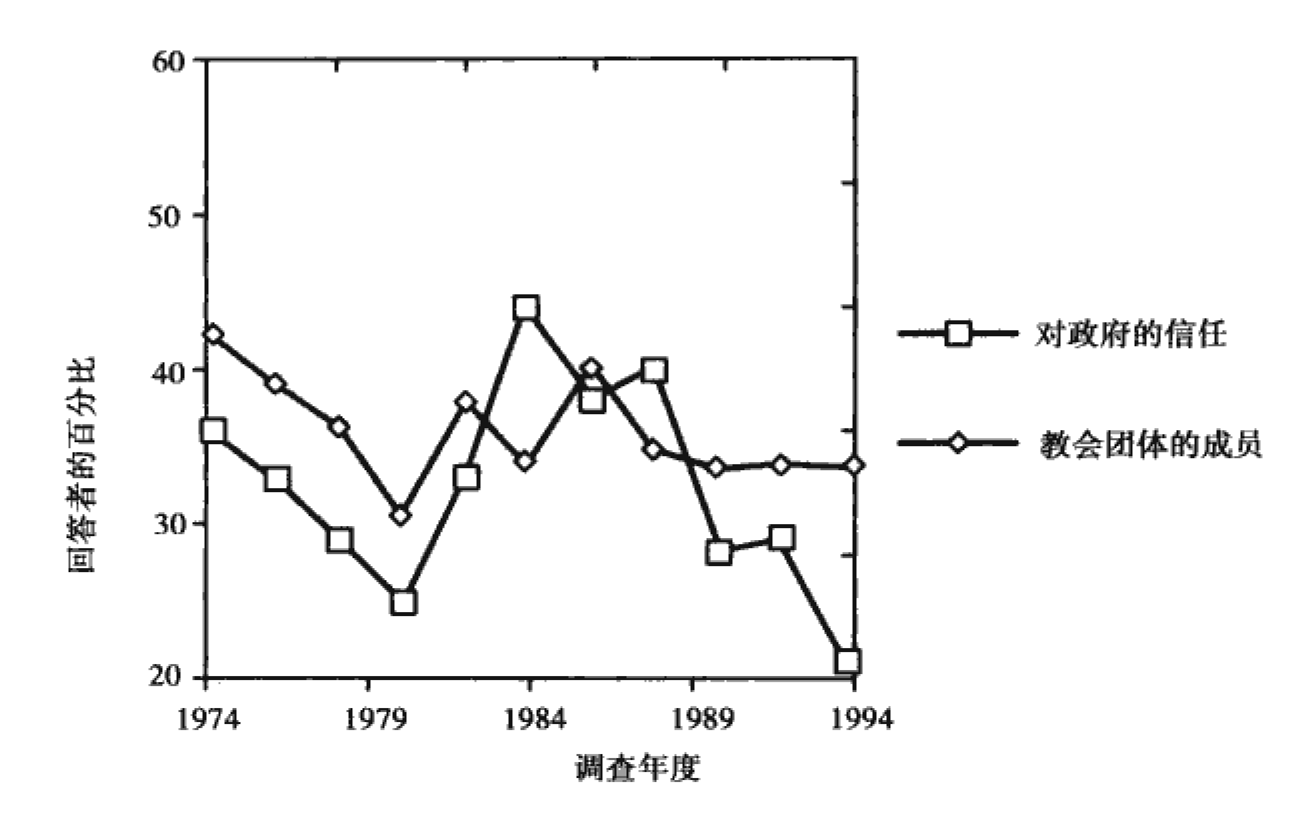


图 3.1 1974—1994 年信任和团体成员身份的趋势

另一种分析方法是看看其他的重要变量对这些结果有何影响。我们对社会资本理论的讨论表明,社团的成员身份一度减少,同时还表明,近来的几代美国人可能不那么愿意参加教会团体、服务俱乐部、友爱组织以及类似的社团。

表 3.3 的数据似乎为此提供了支持。该表显示了不同年龄群体的百分比:其中有些人没参加任何社团,有的加入了一个或两个团体,有的加入了三个或更多团体。由于我们感兴趣的不是年龄而是代,因此我们先圈定一个晋级为成人的时间,根据该时间确定他们大致的出生年代。[⑩]正如我们能看到的,在 20 世纪 50 年代变为成人的回答者("1925—1934 年段"和"1935—1944 年段")最愿意声称他们自己是某个社团的成员。事实上,其中每个年龄段的人中,都有 31% 的人加入了三个或更多团体。相反,到 20 世纪 90 年代变为成人的人("1965—1974 年段")最不大可能报告他们参加了多个组织。在我们设定用来测量社会资本的变量上,他们得分最低。

表 3.3 依年龄段划分的社团成员身份水平

	年龄群				
出生年段	1925—1934	1935—1944	1945—1954	1955—1964	1965—1974
进入 18 岁	1948	1958	1968	1978	1988
不属于任何团体	25%	26%	29%	36%	37%
属于一到两个社团	44%	42%	42%	39%	41%
属于三个以上社团	31%	31%	29%	25%	22%

来源:作者依据 1972—2000 年 GSS 数据库中累积的原始数据计算得到的结果。

⑩ 该表提供的是取自许多 GSS 年度民意调查的"综合"数据。这意味着,1972 年调查时,一个 20 岁的回答者就放在"1945—1954 年段",而 1992 年调查时,20 岁的回答者则放在"1965-1974 年段"。

我们的结论是，有些事情在发生：信任和某些政治参与形式在下降，美国年轻一代的社会资本水平最低。这与普特南的理论是一致的，但并没有“证实”它。我们还可以做更多的工作，来检验这些东西是如何关联在一起的；不过，与开始的时候相比，我们现在对这个问题有了更好的理解。

要强调的是，通过参考多重现实检验，你就能了解到你建构的那种现实检验的功效。同时，其他研究能从整体上对你的发现进行核查。

理解结果

评估概念操作和变量测量

既然我们对这个主题有了一些研究经验，那就根据已经发生的事情，回想一下有关策略的每一个步骤。想出一种方法来操作一个概念与使它如研究过程中所预期的那样起作用，这两者之间有很大的区别。

这项研究计划有一些问题：

1. 你能确定政治参与的测量方法是有效的吗？
2. 我们用作社会资本指标的团体，是否在建构人们的社会关系上扮演相同的角色？社会资本仅仅是研究者为解释数据而创造的一种抽象概念吗？
3. 我们如何能确定社团的成员身份与政治参与之间存在着因果关系？确定团体的成员身份和对政府的信任之间存在着因果关系？因果关系的方向是怎么样的？
4. 关于社会资本的争论是美国独有的吗？我们能否在欧洲或其他地方的调查数据中发现相似的模型？

我们还需要考虑一下，调查数据会以什么样的方式出现偏差。个人在回答问题时的精神状态是个问题。各种各样的因素都能影响到答案。例如：女性访谈者从女性那里获得的答案可能不同于男性访谈者的。在不同的时间和地点重复研究很有用。如果你要大学生填写调查问卷，就得准备好应付校园里的小聪明。爱开玩笑、闪烁其辞和故意作对，人们能用许多方式搞糟调查问卷。

另外一种可能阻碍研究者获得真实答案的因素是，回答者对研究计划、研究者、问题或答案的保密性存有怀疑。以前有位大二学生就大麻的个人用途对一个教会学院的职工做过调查。她私下做的访谈，并向每个员工保证答案是保密的——数据总结会按照部门和级别打散，最终的报告会收藏在图书馆。

然而,可想而知,一个小部门里的低级职工会对访谈者承诺的保密性保持警惕,因此在回答这类可能遭致检举甚至迫害的问题时不那么诚实。

在和人打交道时,科学代替不了机智。

测量方法得当吗

自我批评不是很受欢迎的事,但是在社会科学中,它能满足两个特殊的目的。它显然有助于你在完成计划后重新检查它,确保你每步都做得足够好,可以直接得出结论。不过,重新检查还有另一项功能。在处理如社会现象测量这类难以捉摸的事情时,在提出和使用测量方法的过程中学到的任何东西,都需要共享。研究开始的时候,测量方法可能看起来令人印象深刻。然而,实际中用起来时,可能会发现一些意想不到的弱点,如果在结果中把这些弱点作为一部分陈述出来,就能节省其他人的许多工作。

在罗伯特·普特南的社会资本研究计划中,人们参加了成百上千的不同社团。在各种著作中,普特南强调,重要的是团体间的不同混合——而教堂、体育和艺术俱乐部通常是作为社会资本的基本储存库而被提及。[11] 我们决定要测量的那些特殊团体之所以重要,是因为不同的测量方法通常会导致不同的结果。尽管图3.1表明教会团体的减少对应着信任的降低,但同样的GSS调查问题却表明其他社团的成员在增加(体育团体、学校的友爱组织和职业协会)。

那么,究竟是哪些团体最好地孕育出民主社会必须的规范、网络和社会关系呢?我们需要了解这一点,以便了解该使用什么样的方法来测量团体成员身份。如果像保龄球联盟这样的体育团体很重要,我们就可以问别人,他们是否以联盟成员的身份去打保龄球。但是对该问题的回答能如实地测量他们是否或者如何与保龄球同伙打交道吗?此外,如果许多回答"是"的人,仅仅是因为体育教育而上了高中保龄球班,那我们的测量方法还有什么作用?[12]

记住,社会资本背后蕴含的观念是:团体成员身份能促进人与人之间的联系,从而建立起信任。更可信地测量社团影响的方法,可能是询问人们所属团体的信息,然后问他们会花多少时间与该团体中的成员见面。这也许能更直接地测量社团对人们的影响。GSS还包括一些问题,如志愿者的工作情况、参加某个团体活动的时间。

---

⑪ 见 Robert Putnam, *Making Democracy Work: Civic Traditions in Modern Italy* (Princeton, NJ: Princeton University Press, 1993),以及 Robert Putnam, *Bowling Alone: The Collapse and Revival of American Community* (New York: Simon and Schuster, 2000)。

⑫ 导演迈克尔·摩尔(Michael Moore)提到,谋杀多人的凶手埃里克·哈里斯(Eric Harris)和蒂纳·克莱伯德(Dylan Klebold)都在科伦拜恩高中学习保龄球。在戳穿那些有问题的因果推理时,他讽刺地表示,可能造成屠杀的是打保龄球而非社会信任。

那么,这种批评就等于宣布了测量方法是无效的吗？不,不存在完美的测量方法。关键是要能够在可能的候选测量方法面前,保住你自己的测量方法。研究者必须有好的证据支持其变量的测量方法和变量的定义。比较不同测量方法和不同观察手段所获得的发现,有助于建立理解。

统计数字可靠吗

漫长的历史中,五花八门的神秘主义者观察鸟的内脏,以图发现未来的预兆或预测未来。精于统计数字批判的人可能是这个职业的现代继承者(尤其是那些擅长在任何给定的统计数字中发现好消息和坏消息的人)。统计数字本身并不能为社会调查问题提供精确的答案,这个事实使某些人感到震惊,而对另一些人来说是种宽慰。我们可以轻松地说,统计数字会撒谎,或者说它们从不完整地传达整体信息,因此是无用的。但问题是:统计数字(或测量)是对比什么而言的？对比如"多"、"少"、"整个"或"一点"这样的语言概念,确实统计数字可以做到更加精确。虽然带数字的证据可能会误导人,但是文字更能误导人。象征性的暗示、含沙射影的用词、不精确的语言,都能歪曲知识。科学观察方法的优点是,偏见更容易被暴露出来,因为对意义和程序的规定都非常明晰,能被复制。

当然,错误的统计数字就如错误的文字一样,都容易被派上用场,而且科学代替不了常识。随着对统计数字的深入了解,你会发现,研究者通常是用几种统计数字来总结一种状况,而不是只依赖一个指标,以此来弥补任何特定的统计失误。

你的发现与有关领域的理论吻合得如何

尽管一次简单的实验或调查就可能解决你心中的某个疑问,但它同样也能以一种有趣的方式关联到相关主题的理论中所包含的更普遍问题。例如:了解教育如何影响到人们对总统候选人的评价,还是饶有趣味的。但是把那个发现嵌入到与人类境况有关的一整套观念中,则更有趣。在选举领导人时,我们确实值得信任吗？民主确实起作用吗？这些都是大的理论视角,但好理论不一定非要是重大的。有一些不那么普遍的理论却说明了事件的关键片断。

科尔曼、普特南和其他考察社会资本影响的人在启动其研究计划时,讨论了社会理论、民主理论以及社团与政治关系方面的先行研究。他们引用了其他作家的观点,包括大力宣扬社团的"市民"效应的托克维尔。托克维尔本人在19世纪30年代写到,美国民主最重要的一个体现是广泛地参与志愿联盟——提供公共集会地的社团,其中人们学习一些技能,让自己成为更好的公民。这种讨论突显了与普特南的研究的相关性。他们表明,对政治失去兴趣

的原因可能是深层的社会变革,而不是当今的一群政治家,或是现今对负面广告的鄙视。

在评价本章提出的研究时,参考一下你做过的广泛阅读。另外,如果时间允许,多调查一下其他人关于社会资本(或社团成员身份)与政治参与之关系的发现。想一想,如果在诸如慈善会、社区组织、运动俱乐部和民间团体这样的各类社会场景中,更多的人彼此交往,民主会怎样得到加强(或能否得到加强)。

科学程序本身是沉闷的。在富有想象力的研究者手中,它变成一个非常有用的工具,但是心灵这个仪器比任何调查程序要微妙得多。当科学成为方法,就被作为科学家的人类掌控。

一位著名的科学家曾经评论到,"科学就是观察",他的意思是说,只关注实验组和控制组的细节、统计数字等,就会遮蔽科学研究的目的:用你的头脑来理解正在发生的事情。[13] 关于一种特定的现象,不存在完美的实验来说明所有的一切。提防那些声称他们已经证实一些东西的人——尤其是用基于数字统计的"事实"所做的证实。把科学方法当做批判工具和发现方法加以使用。挑出易受责难的假设和证据的局限性,这样你就能知道什么被证明了,什么还没有被证明。[14]

把你的工作和理论联系到一起并思考其深远的结果,你可以发挥想象力、创造力,尽管不能漫无边际或沦为空想。有人观察到,在公社中人们能工作得更好、过得更幸福。法国社会学家查尔斯·傅立叶(Charles Fourier)把这一观察拓展成一种观念:人类理解的提高会改进(通过数百年)世界的进程,使其达到乌托邦的境界,人们列队开始每天的生活,海洋将会变成柠檬汽水,友好的鲸鱼带领我们跨洋越海。这可就有点过了。

引入的概念

- 事实(fact)
- 真理(truth)
- 可证伪的陈述(falsifiable statements)
- 描述性研究(descriptive study)
- 关系分析(relational aralysis)
- 操作化(operationalization)
- 社会资本(social capital)

---

[13] Robert Hodes, "Aims and Methods of Scientific Research," Occasional paper no. 9 (New York: American Institute of Marxist Studies, 1968), pp. 11-14.

[14] 对科学使用成问题的例子,见 Daisie Radner & Michael Radner, *Science and Unreason* (Belmont, Calif.: Wadsworth, 1982)。

讨 论

1. 科学中有种“现实检验”牵涉到一种对照：把某个被观察到的关系，同如果变量之间不存在关系会出现怎样的结果加以对照。看本章中的表 3.3：如果年龄段和一个人参加的团体数之间没有关系，那么可能出现什么样的数据？

2. 本章列举的社团是否反映了人们在其中发展“那些推动互利合作的网络、规范和社会信任”的地点？当人们隶属这样的团体时，你想想他们会做什么；他们会同意你个人的看法吗？联盟或职业协会是否也是可以建立社会资本的团体？

3. 你能想到用其他的方法来操作化概念“政治参与”吗？操作化概念“对政府的信任”？

4. 就政治参与问题，你认为回答者能给出即便不完全精确但被社会认可的回答吗？是否有某种方法来设计一项研究，可以避免有效性这个问题？

有用的网址

- 大选研究(http://www.umich.edu/~nes/)

查看一下政治参与和对政府的态度的趋势。该网站有一些表格，它们根据各种人口统计特征把调查问题制成交叉分析表。

- 综合社会调查（http://www.icpsr.umich.edu:8080/GSS/homepage.htm）

查看一下与社会和政治态度有关的数百个问题的回答。该网站允许你在在线 GSS 译码簿中把任何问题的回答制作成交叉分析表。

- Bowlingalone.com(http://www.bowlingalone.com)

罗伯特·普特南著作的升级材料。包括社会资本的定义和书中所用的数据。

第4章 提 炼

Refinements

"如果人们要探索任何新观念的所有可能性，而且相信以后的经验可以用于确证或修正最初的主张，那么热情和深信必不可少——因为科学的繁荣乃是建基于沉思的自由和无情的批判这一双重程序之上的。"

—— S. 图尔敏和 J. 古德费尔德(Stephen Toulmin and June Goodfield)

学着了解社会科学的方法论，就像学玩台球。它们各自的基本元素很简单——在台球中，有球袋的桌子、一些台球和一根球杆；在社会科学中，就是变量、测量和假说。到现在为止，我们关注的一直是简单的击球：一条假说，带有两个相当明显的变量和对变量关系的测量。科学就像台球一样，更精巧的策略都是从基本的技巧变化而来。

优秀的台球选手若非绝对必要，绝不尝试更困难的一击；社会科学家也是如此。同样，为了把错误降到最低并绕过障碍，两个领域中的专家都必须发明新的技巧。我们将在本章和下一章中，以一种稍微不同于前几章的顺序讨论社会科学的元素：先是假说、变量，然后是测量；我们还会对每个元素做些提炼。换句话说，我们将展示的是擦板球入角袋。

假 说

假说并非从不受思想和偏好复杂影响的理智中一次成型的。就像其他人类行为的产物一样，假说是意向、学识和关注所构成的综合体的一个部分。社会科学家就如何处理这个现实，激烈地争论了很长时间。有些人提出，为了维持专业社会科学的名声，集中于"客观"的探索工作，研究者应该尽可能忘掉价值和其他偏见。另一些人坚持认为，忽视假说的来源是徒劳的，因为这会导致研究者忽视他/她自己在处理数据时的一些基本因素。

至于假说如何与理论、模型或范式这样的思维结构相吻合，存在一系列的问题。毕竟，工作的最终目标是形成有用的理论。因此，理论和研究之间的关系需要探索。

最后，在能被吸收进假说的那些关系中，还存在一些更平常的但有影响力的东西。这三个主题——在假说形成中的价值、理论和关系角色，会依次得到讨论。

价值和假说形成

价值这个概念本身很特殊。写作的人经常试图把握价值是什么，如何把一个价值与其他价值隔离开来。麻烦的是，价值很难被孤立出来。我有信仰的自由，但没有消除平等的自由，也没有自由采取某些类型的行为，如偷窃。价值发生在彼此互相调节的境况之网中。[①] 我们所有人体验到的那个矛盾困惑的自我，可能经常会在不同的时刻展现出不同的价值集，只有在一段较长的时期，才出现较大的、明显的价值模型。不过每个人的性格还是有种一致性——足以使我们确实能够对人们的人生定位做出一般评价。

社会科学家经常建议，研究者在提交项目报告时要讨论其价值定位，以此来解决价值与研究的关系问题。由于在形成假说、挑选测量方法和评估结论的每一步中，价值都是一个密不可分的部分，因此这个要求相当不错。不过，详细澄清其价值不能是事后的补记。研究伊始，价值的角色就必须直接面对，否则你就可能看不出价值对研究的影响。例如，有很强宗教信仰的人在研究约会习惯时，提出的问题可能就预设了婚前性交是不道德的。所用的问题很可能反映出这种偏见，诱使回答者谴责他们实际上赞同的行为。

理论、模型和范式

假说或调查现象的理论和模型之间的关系，似乎是清清楚楚的，但是当作者试图把这种关系见诸笔端时，却变得越来越复杂。我们知道，理论是一套有关系的命题集，它试图说明或有时是预测一系列的事件。我们也已讨论过假说是什么。大致说，理论是被某种逻辑框架关联到一起的假说构成的集合。至于理论提供的理解是否有效、是否正确，“理论”这个词语暗含着它在这个问题上有一定程度的不确定性。因此，理论是尝试性的表述。被证明难以被证伪的理论通常包含在“规律”或公理集中。

另外两个词也被列入了讨论范围。科学家使用模型这个词语传达的含意是理论中更严格的秩序和体系。模型提供了对现实的简化，能接受对关键关

---

① 就价值如何与概念和意识形态相关联这个问题，有很精彩的讨论，参见 Michael Freeden, *Ideologies and Political Theory: A Conceptual Approach*(England: Oxford University Press, 1996), Part One。



“哦,葛瑞森先生,您是我听到的第一个在现实生活中使用‘范式’这个词的人!”

系的检查。例如,经济学家竭力构造理论模型,这样,失业、通货膨胀和其他与经济行为有关的主要变量就在数学上被关联起来了。

范式(源自拉丁语词根,意思是“模型”)这个词语指更大的理解框架,由更大的科学家共同体共享,它把小尺度的理论和研究组织在一起。在古代的数个世代中,统治天文学的范式认为地球是宇宙的中心。早期的宇宙观察者试图在这个背景下说明所有其他的恒星现象。当然,最终由于出现了有力得多的解释,所以这个范式轰然倒塌了。

社会科学中很少有**规律**和**公理**,有一些普遍性的范式、许多理论,以及近来还有些有趣的模型。对于做社会科学调查的新手来说,他们最好把理论看做是一种研究指导——一种组织和有效利用洞见的方法,可以用于去粗取精。

社会科学中,理论有两种一般的工作模式:归纳的和演绎的。**归纳**是指通过积累和总结各种研究以建立理论;演绎牵涉到运用一种理论的逻辑产生一些能被检验的命题。

科学最广泛的形象是,研究者通过逐步的调查过程收集信息单元,然后把它们建成理论。

接下来就检验该理论能否说明某个现象。把科学仅仅看成归纳,接受这种看法的危害在于,在建立研究的过程中使用的范畴可能反映了一种隐含的理论。结果,所谓的归纳发现是一种隐蔽形式的演绎。科学程序要求理论命题以一种可证伪的形式表达,也就是说,它们能接受观察的检验,以减少这样的偏差。尽管这个要求看似清清楚楚,但社会调查的价值负载沉重,而且现实检验工具有限,因此很容易出现错误的判断。

在关联理论与研究方面,演绎是一种越来越普遍的方法。由于批评家对社会科学所谓的"客观"本质的抨击,研究者在此压力下开始理解到,演绎不知不觉地渗透到假说通常使用的基本概念的形成中。在美国文化中,资本主义政治经济体制的普遍境况,导致许多政治科学家、社会学家和经济学家把我们的体制当做好社会的规范,并且把所有非市场的行为模型归结到脱轨的、有负面效果的、不发达的等这类负面范畴中。实质上,这些标签的内涵是从一个更大的理论中演绎出来的,该理论暗含了政治经济体制的本质和正确性。不过,这些标签却以被归纳确定的"科学的"名号出现。②

从这样受文化约束的假定出发,人们就更容易论证说,如果某个人依照物质私利之外的动机采取行为,那么他就是"没适应的"或"非理性的",或者需要治疗、调节。事实上,被标上非理性的行为可以满足资本主义社会中被压抑的需要,因而有助于人们适应困难的环境——例如,一位穷人买辆豪华轿车的行为。拥有一辆轿车,是这位穷人展示成功、求得自重和关注的一种方式。这种耗尽其食物预算的支出,可能令中等阶层的观察者深感愚蠢,这主要是因为,那些好歹有工作的中等阶层观察者不会遭受连续的拒绝和个人屈辱所带来的压力。

由于演绎是一种自然的思维模式,所以需要用于科学探索。从理论出发的演绎通常提供了研究领域的基本议程。已确立的理论能指导人们找到许多特殊谜题的解决办法。在从事新一轮的说明任务时,演绎法是相当值得尝试的。

我们不需要过于关注归纳和演绎。优秀的科学研究总是包含一些元素,它们能让其他持不同视角的人判断其价值。记住这些要点的主要原因,是要警惕自欺、警惕其他人在提出其科学发现时的一些误导性的方式。

远在你有能力提出理论以前,你多半就已经是其他人兜售的理论的消费者。理论赋予结果以意义,在运用研究结果时,关于理论有个预防性的问题需要提出。它类似于研究背后的价值这个问题,就是对理论视角的理解,而理论视角是研究得以实施的前提。在读社会科学著作时,一定要仔细关注引言和前言——这里通常点出了作者工作的关键之处。

与此同时,不要害怕运用那些理论性的说明,要把它们作为你自己努力的一个指南。科学是民主的,任何人都能对一个理论做研究批判,或者尝试用新的方式加以拓展。留意一个领域中占支配地位的理论,借用其观点来看看能为一种现象提供什么样的说明,这能节省你自己的时间。

---

② 参见 Murray Edelman, *Political Languages: Words That Succeed and Policies That Fail*(New York: Academic Press, 1977)。

假说中的关系

自变量和因变量。并不是所有的变量都是同等的。如果社会科学只是要表明,偏见与无知有关,年轻与叛逆有关,智商与哺乳有关,那么社会科学家并没有做到如我们合理期待的那样多。人们的偏见是因为他们无知吗?或者说无知是因为他们被偏见蒙蔽吗?哪个是先导?我们几乎脱口而出谁因谁果,但没有,这是因为对因果关系的判决性证明需要精密的程序。变量中的自变量和因变量这对概念是一种方法,可以在不用通盘解决的情况下,巧妙地处理因果问题。

自变量影响因变量。例如,热量增加,空气中就能保存更多的水分。热量是自变量,而能悬浮在空气中的水分是因变量。水分的状况取决于温度的变化。如果空气潮湿而热量减少,水就会从空气中滴落——这就是社会科学家所谓的"降雨"。

在第3章的例子中,普特南和其他人认为,政治参与取决于人们是否积极参与自愿社团。政治参与是因变量,社团中的成员身份是自变量。据说,积极投入社团活动能建立个人的关系和信任,使人们有能力参与政治活动。这样就有了下述假说:社团活动越多,政治参与就越多。

有个好办法可以让你看出一种被假定的依赖关系是否有意义:颠倒一下你正考虑的那种关系。投票或参加政治活动是否能导致人们加入教会团体或保龄球联盟?也许能,但是你需要一个有力的理论来说明,为什么这种关系会在那个方向发生作用。社会资本理论就提供了一种令人信服的论证让我们相信,社会活动是政治活动的先导。

在很长的一段时间里,自变量和因变量这两个概念没什么麻烦之处。看看教育和投票决定之间的关系,就能轻易发现,投票不可能是教育的原因。但有个事实却颇为麻烦:自变量和因变量的关系在令人信服地得到证明前,只是研究者的一种想象虚构。自变量和因变量之间的关系是研究者假说性地提出的:他们发明了这些关系,然后用观察和分析来检验它们,看它们是否确实如想象的那样发挥作用。

当我们用社会科学家最喜爱的形式——表格——来考察时,自变量和因变量的问题就变得更容易理解了。表格是展示数据的一种方法,但表格背后通常隐藏着新手容易忽视的假说。

看看表4.1。哪个是自变量?哪个是因变量?你如何重构这些数据所支持的假说?

表4.1 富人与穷人的政治活动:收入群体的活动比例(%)

	年收入					总计
	低于 $15 000	$15～35 000	$35～75 000	75～125 000	高于 $125 000	
用于竞选活动的钱	2	14	29	20	35	100
选举中的投票	14	32	43	7	4	100
接触官员	12	28	46	8	6	100
抗议	12	30	46	5	7	100
投入活动的时间	13	30	39	8	10	100
占美国人口的比例	19	36	36	6	3	100

来源:Sidney Verba, Kay Lehman Schlozmen & Henry E. Brady, "The Big Tilt: Participatory Inequality in American," *The American Prospect*(May/ June 1997),可见 http://www.prospect.org/archives /32/32verbfs.html。

两个变量是收入和政治活动。这些数据表明了哪些收入团体主导了某些活动。它们揭示了这两个变量之间的何种关系?答案是:最富有的群体尽管只占美国人口的3%,但比起不那么富有的群体而言,他们参与各类政治活动的比例要大得多。他们为竞选提供了35%的资金支持。最穷困的群体占19%的人口,但在投票人赞助选举资金中只占2%的份额。因此,收入影响了各类人群的政治参与。

因此,收入被认为是自变量,而政治活动是因变量。要检查"自变量"和"因变量"名称的分派,颠倒一下假说。比如说,投票行为决定了收入?胡说八道!

表4.1展示的是通常情况下表格出现的形式。自变量横列于表格顶端,因变量沿侧边朝下放置。表格以这种标准的方式呈现,研究者想都不用想就可以定位关系。不过,看着表格,阐述它旨在检验的假说,是一种很有益的实践。作者可能出于强调、风格或方便,颠倒因变量和自变量的惯常定位。

替代变量、前置变量和中介变量。在提出有说服力的假说时,核心问题之一是理解变量之间的关系。在就变量间的关系提出假说的过程中,你需要留意你没有选择的变量,它们可能影响到关系的变化。社会科学家通常会提到替代变量、前置变量和中介变量。

三个术语都有常识意义。替代变量是额外的自变量,它影响到因变量的变化。前置变量是某个先行物。例如,出生之前要先怀孕。中介意味着中途出现。我们会更精确地阐述这些概念。

如果有人考虑一些影响人们对政治活动的贡献的变量,就会有这几个出现:性别、种族、职业、党派、对政府的态度,其他一些因素也可能有影响。这些都是替代变量。不过,在收入和对政治活动的贡献之间建立关联(见表4.1)是有用的,尽管对人们为何做贡献或不做贡献的完整说明需要涵盖所有重要

的替代变量的影响。收入确实影响到贡献,但性别和其他变量也介入了。如果重要的变量被忽略了,那么结果可能是无意义的,或者就如社会科学家喜欢说的,是虚假的。我们将在下一章再讨论虚假性问题。

对前置变量的一种经典阐述来自对投票行为的研究历史。早期的调查清楚表明,教育层次越高的人越倾向于投票给共和党。从这种关系我们可以推测,受过良好教育的人在政治上比较保守。然而,结果证明,有个强有力的前置变量影响到了教育水平和投票行为,它就是父母的财富。事实上,受过高等教育的人一般来自更富的家庭,而更富的家庭更有可能投票给共和党。在教育和投票行为的关联中,被测量的实质上是父母的财富对小孩政治倾向的先行影响。

再来看中介变量。假如别人告诉你,每片 Hollygood 面包所含的卡路里比其他 6 个品牌的低。这个广告导致你推断:自变量是 Hollygood 低卡路里的特殊配方。但结果却发现,Hollygood 公司的面包片比其他牌子的薄。面包中含有的卡路里量实质上与 Sunshaft 面包甚至 Wondergoo 面包的差不多。面包切片的薄度,就是面包品质和每片面包的卡路里含量之间的中介变量。

还有一个更好的例子:考虑一下教育和社会地位之间的关系,这两个变量有正相关关系。然而,每个人都认识教育程度一般但地位很高的人。其原因可能是有另一个变量渗入了:事业成功。

要了解事业成功是如何介入教育与地位之间的,想想你认识的那些教育程度低但由于事业成功而地位不错的人(A 组)。再想想教育良好、取得成功且地位较高的人(B 组)。还想想教育良好而事业不顺且从传统标准来看地位一般的人(C 组)。

如果你仅仅关注教育与地位之间的关系,或者是事业与地位之间的关系,而非所有的三个变量,那么你可能会错过任两个变量之间的关系。A 组使你认为,教育和地位之间没什么关联;而 B 组则表明,地位和教育间的关系就如同花生糊与花生酱之间的关系。而 C 组的人尽管跟 B 组的一样受到良好的教育,但地位只是一般。如果只考虑事业成功与地位之间的关系,也会产生同样的混乱。

一般来说,受到良好教育的人(B 组)比教育程度低的人(A 组)地位要高。因此我们可以表明,教育有助于成功。不过,事业成功会介入到教育和常规的社会地位之中。

为了避免陷入替代变量、前置变量和中介变量的困惑中,我们可以在提出假说前做些思考。选取了因变量后,问问自己所有可能的自变量会有哪些。如果你想要说明为什么某些人是宿命论者,就想想可能影响这种心灵状态的所有变量。可能因素包括他们的工作性质、财务问题、单恋、深层性格、气候或同伴的影响。事实上,大部分乃至所有的社会现象都受到数个变量的影响。

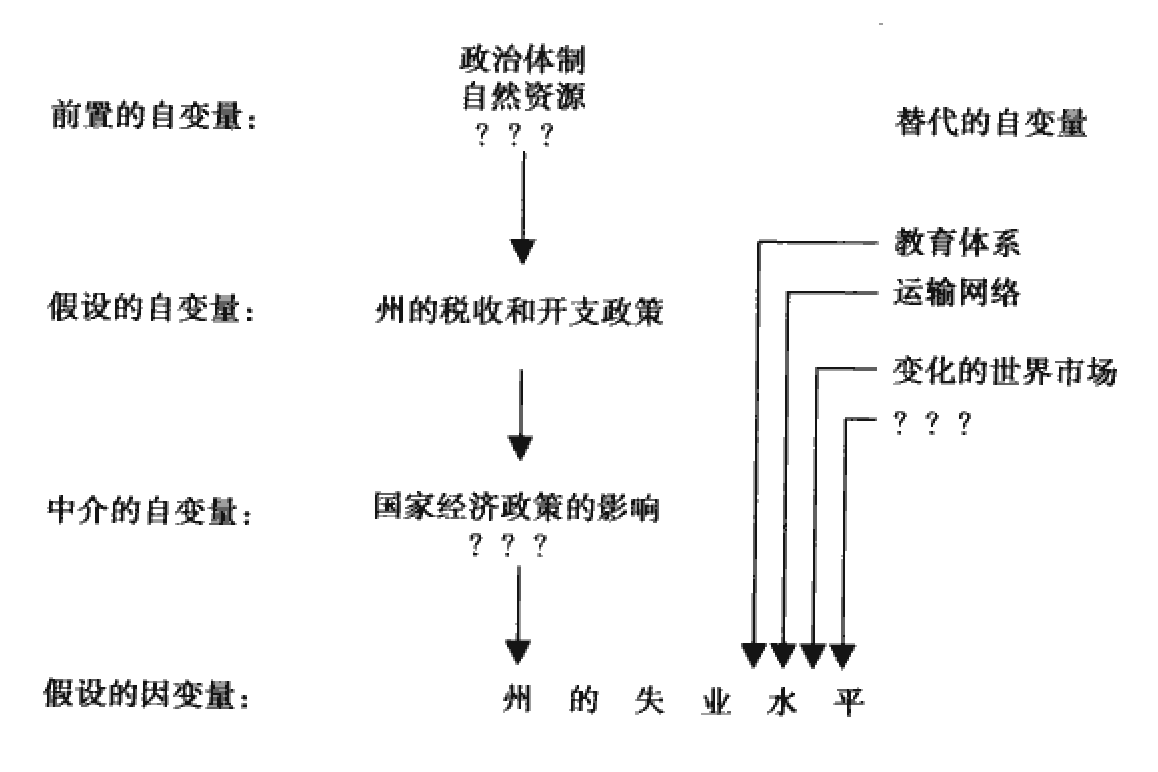


图 4.1 影响失业的自变量

提出对替代变量、前置变量和中介变量的担忧，不是要打击你开展你感兴趣的研究的积极性，而是让你记住它们，不至于把相关性混同于因果关系。

再举个例子，考虑竞选活动中频繁听到的一个争论：州税对就业率的影响。批评政府支出的人争论说，降低税收和开支会吸引不喜欢纳税的企业，从而增加新的工作和降低失业，这样就会刺激该州的经济。在图 4.1 中，我们指出了可能对州的失业率造成影响的前置、中介和替代变量。

仔细地梳理能揭示许多卷入的自变量，其中任何一个都可能比我们假设的变量更重要，也就是比州的税收和开支政策更重要。事实上，这些变量之间的关系相当复杂。州的政治制度，如直接民主制，能利用立法提案和公民表决来降低税收和开支。然而，只要较高的州税为教育体系提供了资金支持，就可能成为促进该州就业的关键。③

一旦你认识了对因变量有重大影响的那些变量，就有办法把某个变量的影响与其他的分离出来。最简单的技巧是“控制”：保持一个变量不变，检验其他两个变量的相互关系。在父母财富、教育和投票行为之关系的这个例子中，你可以挑选一个样本，其中回答者有不同的教育层次，来自不同的家境。假如结果表明，出自富有家庭且教育程度高的孩子主要倾向于共和党，而家境不好而教育程度高的孩子主要倾向于民主党，你就知道，在投票行为上，教育的影响远不如家境的影响。表 4.2 和表 4.3 说明了这个结果。

③ 例如，参见 Bryan D. Jones, “Public Policies and Economic Growth in the American States,” *Journal of Politics* 52 (1990): 219-234。Jone 发现，公共部门的整体规模与经济衰退之间没有关联，用于教育、高速公路、警务和消防的开支，与就业增加、经济增长有关联。

随着你在方法论上的经验和熟练度的增长，你将发现许多技术可以用于梳理这些关系。要处理变量梳理这个问题，第一步是理解要建入假说的那些不同层次的关系。

假说中的关系层次

相对最常见的语句而言，假说最明显的特征就是它在阐述每个术语时的谨慎。我们已经看到，变量的选择本身是个严肃的任务；同样，变量之间被阐述的关系也是如此。为了稍微延伸你的想象力，系统地考虑两个或更多变量之间可能体现出的关系，是很值得的。它们构成了一个连续统，我们将简要地讨论一下表4.4中的每种关系。

第一种关系，虚无假说，真是天才的创造。记住，假说是想象的关系，然后投入检验。首先断定无关，然后检验，看看虚无假说是不是能被否证，也就是说，看是否能证明确实存在某种关系。这里面还有文章可做。

表4.2 教育与政党认同

政党认同	低教育程度	高教育程度	总计
民主党	150	50	200
共和党	50	150	200
总计	200	200	400

来源：模拟的。

表4.3 教育与政党认同，控制财富不变

政党认同	来自贫困家庭			来自富裕家庭		
	低教育程度	高教育程度	总计	低教育程度	高教育程度	总计
民主党	100	45	145	50	5	55
共和党	5	10	15	45	140	185
小计			160			240

来源：模拟的。

表4.4 变量关系的类型

关 系	意 义
虚无的	认定没有什么关系存在。
推断的/关联的	认定有一种关系，但它处理的是一个变量对另一个变量的影响程度。
正向的/反向的	认定有种特殊的关联性关系，其中两个变量之间存在一种可预测的关联——或者一个变量随另一个变量的增加而递增（正向），或者一个变量增加而另一个变量减少（反向）。
因果的	认定一个变量的变化是另一个变量变化的结果。

虚无假说的功用在于不让你做臆断——不是在提前为你阐述的关系做辩护。除了不承诺特定的关系外，你还保留了一种可能性：有种更实质的关系可能刻画了变量之间的关联。也许情况是，有一种推断的或关联的关系，将会产生于现实检验。甚至可能存在一种正向的或反向的关系，但这些可能性还要等到检验的时候出现。

虚无假说正好响应了谨慎的社会调查策略的要求。要否证虚无假说，只能通过证明两个变量之间存在着某种关联。因果关系需要大量的证据，它与虚无假说相对，落在连续统的另一端。

推断的和关联的关系可以得到检验，作为向因果关系迈进的准备。更次要的关系尽管有趣，也只是一种筛选手段。在教育和候选人对投票人的评估之关系一例中，如果可以证明统计上重要的关联，那么我们就有理由向前推进，分离出那些可能影响到此关联但旁枝末节的错误源头。一旦完成，我们就可以检验因果关系的替代源头，看看因果假说能否得到辩护。

关于因果关系，有些事情需要理解。首先，社会科学的最终目标可能是确定孰因孰果。因此，确定因果性非常有趣。其次，它也是最难处理的关系，因为它需要最高程度的证明。要证明 A 是 B 的原因，你需要表明：

1. **A 先于 B 发生**，显而易见，不是吗？
2. **A 的发生与 B 的发生有关**。这同样也是显而易见的，但事件间的关联并不总是容易发现的。例如，有些历史学家发现，中世纪改革者的饮食同他们细腻的洞察力之间，存在着稳定的关联。据称，圣女贞德误食食物，弄坏了消化系统，因而成为一位有远见卓识却昙花一现的女英雄！
3. **A 导致 B；不存在其他的变量（C）能消除 A 给 B 带来的变化**。这是难关所在。在一种境况中，只留下一个而消除所有其他的可能影响，总是非常困难的事。实验社会科学有一门历史悠久的技术：挑选两组受试，尽可能使两组的环境保持一致，然后在一组中引入被怀疑的因果变量（实验组），而另一组不引入（控制组）。

至于实验-控制组技术的运用中出现的问题，有个典型的例子，就是“霍桑”实验，其中，作为实验组的工人被安置在更愉悦的物理环境中做装配线工作。该实验组的产量一直高于在普通工场条件中工作的控制组。问题是，后来发现，增加的产量主要是另一个变量——经理和实验者本人对实验组的特殊关注的结果，而非物理环境。实验者不知不觉地引入了一些没有被控制的心理因素：两组的差别不仅仅是物理环境上的，因此破坏了实验-控制组程序，而导致结果的无效。

在大多数社会科学家看来，对因果关系的理解是假说形成和检验这一长串过程的顶点。在开始实施一系列的实验时，常见的做法就是分离出一个变量：它与被引起的事件有最明显的关联。这种方式可以识别那些被怀疑的因

果关系来源。后续的逻辑步骤通常需要非常精心的实验安排。因此，这个领域的新手最好安心于那些更容易处理的关系。

由于社会科学研究影响到一些重要的私人事务，所以很难在假说形成过程中培养谨慎的习惯。大多数新手对自己的假说自视过高，导致他们陷入测量的困难和为得出结论而超负荷工作的失望中。在确定某种待检验的关系的强度时，要考虑一下可用的测量方法和数据资源。一项被完整报导的研究实验，总是包含着研究者对结论的深远后果的思考。但是，如果研究本身能观察到假说陈述和测量技巧的明显限制，那么这些思考就更合乎人意了。

在假说中建立变量之间的逻辑关系和检验这些关系是否成立，当然是两码事。要看一种被假设的关系是否确实得到观察的证实，我们需要转向操作化和测量技术。

变 量

操作概念

早在我们对社会科学概念的讨论中，就看到语言是如何从对不同现象的命名而开始的。对于一种特定的行为，人们专门约定一个名称，正是这种约定，构成了社会科学的语言。操作变量实质上意味着，对一种行为的命名要适合于某种观察和测量该行为的特殊方法。从某种意义上来说，变量操作就是颠倒语言的形成过程：从你感兴趣的现象的名称出发，往后回溯，找到一些方法，把那个名称与它所指称的特殊行为匹配起来。

操作化这个词语使这里讨论的过程听上去既特别又专业，然而事实上，它在日常生活中是老生常谈的。一天晚上，本书的一位作者在酒吧里听到关于肯塔基人和西弗吉尼亚人哪个更好的争论。“讨论”围绕下述观察展开：有个人的堂弟的叔叔的岳父来自肯塔基，他不是个好东西。不过，相比之下，另一个人的前任老板的夫人的侄子来自西弗吉尼亚，他天生就该绞死！几轮争论后，显然，肯塔基人和西弗吉尼亚人的品格这个变量，已经根据生活在那两个州的人们的犯罪行为倾向而被操作化了。

在任何科学的发展过程中，用于指称被谨慎指定的对象、时间或行为的变量名称，其数量会逐渐增加。如今在社会科学中，也有整套的变量目录，它们根据特殊的行为和可能的测量得到操作。[4]

---

④ 见 John Robinson, Phillip Shaver & Lawrence Wrightsman, *Measures of Political Attitudes* (San Diego, Calif.: Academic Press, 1999)。

有点值得庆幸的是,你感兴趣的变量早已被各种方式操作化了。即便如此,你也需要了解一些操作技巧,以获得分析的灵活性,同时也对其他人的工作保持批判态度。此外,当变量需要的测量形式超出你的资源范围时,你需要学会避开将出现的问题。有两种方法可用于处理由于某种原因而不易操作化的变量:替代和分解。

假如你的假说是这样的:*教育程度越高的人,越有可能是社会主义者*。教育不难操作化:呆在学校的年头会告诉你接受了多少正式教育。

人们是不是社会主义者,如果是的,他们的社会主义倾向有多强烈,则是另一回事。社会主义这种意识形态集合了一套复杂的理论、历史观、行为计划和好坏标准。如果你意识到,讨论这个主题的学者很难就社会主义的意义达成一致,就知道这套东西有多难理解。分离出社会主义的标准定义是个难题,但难上加难的是,在处理这个词的非共有解释上存在着问题:研究者很可能接受了正式的意识形态概念的训练,而作为调查样本的回答者却可能认为,社会主义者是那些喜欢氟化水的人们。

因此,"你是社会主义者吗?如果是,到了什么程度?"向别人提这样的问题没多大意义。这个问题的答案会产生一些关于自我知觉(self-perception)的有趣资料,但是如果把它作为一种方法,把回答者的态度与某种像社会主义哲学一样精致的东西关联起来,则显得太随便。替代变量*社会主义*,或许可以解决其中一些问题。我们可以找到另一个变量,它能确定更直接卷入的态度,并对之做具体的处理。看看这条假说如何:*人们受教育越多,越支持工人参与管理*。

替代的优势在于,我们可以就大多数人都能提出意见的事务提出问题,而且这种提问方式能把它们与该事务关联起来。即便它是个替代,但它圈定了社会主义意识形态的一个重要元素,从而提供了与那条一般性假说有关的信息。

*分解*是另一种处理复杂变量的方式。行为一般都不简单,它发生在活动、态度和倾向互相关联的语境中。社会科学家所处理的变量,通常可以被看成是行为要素的组合。例如,变量*异化*可以分解成四种具体的特点:失范的、无力的、无意义的和无助的,它们与人们在异化时的通常感受是关联在一起的。态度量表的提出就是为了测量异化的每个态度要素。通过组合所有的四个态度或感受的测量,你获得的资料就可以称得上是与异化有关的。

变量的维度

变量通常有不同的维度。测量人格的心理学家可能会提出内向人格和外向人格的分类。他/她还可能提出在1-10的度量表上对攻击性-回避性做特征描述。它们反映了一个变量的不同维度,此变量即人格。

民意通常用各种维度加以分析：

方向：意见的赞同或反对

位置：在从赞同到反对的度量表上，该意见落在何处？

强度：持该意见有多强烈或多无力？

稳定性：它的可变性如何？

潜在性：它离意见结构的表层有多近？

突出性：与个人所持的其他意见相比，那个意见的重要性如何？⑤

所有这些维度都包括不同的测量可能性，也有各种可用的技术来处理它们。意见的方向只需要清楚告诉我们意见是“对”是“错”。另一方面，突出性可以把意见从不突出到非常突出做顺序排列。意见的强度则暗示了用刻度表示的可能。

在对一个变量进行大规模处理前，好好想想你着眼的是哪个维度，想想还有哪些其他的可能维度。挑选出最有可能触及变量核心的维度。通过考虑替代的变量，你就能选定：哪些维度把握到了变量的关键，哪些维度可以用你资源范围内的方法得到测量。与此同时，对变量的不同维度的理解，可以让你看到，在理解该变量方面，哪些工作已经做了，哪些还没做。

科学最持久的神秘之一，就是它居然能完全等同于测量。就如本章力图说明的，科学真正的创造力在于变量的操作化和假说的设计。它们经常需要真正的创造力。尽管有时候测量接近于艺术形式，但它更多的是一种技术和数学概念的系统应用。就如我们将在下章看到的，测量有自己的逻辑和巧妙设计。

引入的概念

价值(values)

理论(theory)

模型(models)

范式(paradigms)

前置变量(antecedent variable)

中介变量(intervening variable)

虚无假说(null hypothesis)

推断的关系(inferential relationship)

---

⑤ 改自 Bradlee Karan, “Public Opinion and the New Ohio Criminal Code,” The College of Wooster Symposium on Public Opinion and the New Ohio Criminal Code, July 9-30, pp. 6-8，以及 Vladimir Orlando Key Jr., *Public Opinion and American Democracy* (New York: Knopf, 1961), pp. 11-18。

Key 用属性而非维度讨论了变量。在讨论民意时，他在我们用“位置”的地方用了词语“维度”。就近来的用法来看，词语“属性”变成了一个通名，指称一个变量的所有特性：它的测量、它的各种实质成分，或是维度，但它们已经获得了更明确的意义，也就是我们提到的意义。

规律(laws)
公理(axioms)
归纳(introduction)
演绎(deduction)
自变量(independent variable)
因变量(dependent variable)
替代变量(alternative variable)
关联的关系(correlative relationship)
正向关系(direct relationship)
反向关系(inverse relationship)
因果关系(causal relationship)
变量替代(variable sabstitation)
变量分解(variable division)
变量的维度(dimensions of variables)

讨 论

注:检查表4.2和表4.3的数据。在该例中,表4.2考察的是教育和党派认同之间的关系。表4.3考察的是来自贫困和富有两种家庭的人对党派的认同。表4.3引入了对家庭财产的控制,因为党派认同也可能与家庭财产有关。

1. 当我们考察表4.3中的两个团体时,表4.2中的初始模型是如何变化的?

2. 党派认同和教育之间的关系是否受到收入的影响?

3. 关于教育和党派认同之间的关系,表4.3中的结果有何指示?

4. 对于两个表格展示的模型,你能想到有何逻辑说明吗? 哪个(些)是因变量? 哪个(些)是自变量? 为什么?

5. 考虑布洛金顿等人(Brockington et al.)的文章(附录B)的回归分析中的变量,如果认为少数族群的代表性水平取决于选举规则和社团的少数族群,这种看法是否有意义? 该因果过程是否也能在其他方向发生?

第5章 变量和关系的测量

Measuring Variables and Relationships

"为月织衣裳，夜夜勤裁量。"

——纳撒尼尔·沃德(Nathaniel Ward)

科学家主要测量三样东西：变量、与变量相关的数据具有意义的几率，以及变量间的关系。每项测量任务都有独特的途径和统计技术。当我们考察这些任务的完成需要运用各种观念时，请记住：测量的精确性几乎总是名大于实。

本章在相当宽泛的意义上使用测量这个术语。第一个主题是变异的测量，我们将考察适宜于不同类型变量的各种测量。接下来，我们考察一些用于描述数据意义和代表性的技术，这些数据是通过科学程序获得的。有些技术可以相当精确地判断，一组数据是否只是奇特样本的结果，而不是有意义的测量的结果。在此过渡部分，我们将表明该如何构造抽样调查，还会讨论一些常见的民意调查错误。然后，我们就变量间的关系测量提出一些看法。我们的目的是要掌握一些简化数据的基本工具，把与两个或更多相关变量有关的数据简化为一个统计量，使之能刻画变量之间的关系。

按照惯例，测量作为一个术语只适用于第一个主题。第二个主题关注的是数据的意义和代表性问题，并且使用了概率，而概率在最狭隘的意义上并不是一种测量形式。第三个主题通常被认为是刻画事物之关联的问题，严格说来也不是测量的问题。然而，所有这三个主题都与确定某物的数量相关：变异、意义和关联。因此，所有这三个主题一直都被恰当地归在测量这一总类目之下。

测量变量:测量层次

测量是个富有欺骗性的主题。初看起来,它似乎很简单——测量是回答“多少”的问题。当谈论长度或重量时,这个问题看上去相当容易回答,但当考虑信息化水平、人格特征、情感和态度诸如此类常见的社会科学素材时,这个问题就不那么容易回答了。困难的原因,与其说在于总计事物的个数,倒不如说在于被计算的事物的本质。

例如,在测量变量时,有三方面的考虑可以确定我们能尝试哪一测量层次,因而也确定了使用该变量可以提出哪些类型的假说性关系:

1. 变量的属性或特征。
2. 适合于这些属性的测量技术。
3. 根据变量的属性和可资利用的技术,有哪些可能的测量层次。

举个例子,考虑婚姻状况这个变量。它指的是根据法律界定的一种分类:单身的(又可再分为未婚的、离异的或丧偶的)或已婚的(或许又可再分为一夫一妻或非一夫一妻)。在“测量”某人的婚姻状况时,这个变量的属性规定,你只能对它进行分类——不可能说某人结婚的程度非常高,或某人结婚的程度非常低。从法律的角度看,你要么结婚要么没结婚。一旦这样一种属性已定,婚姻状况这个变量并不需要十分花哨的测量技术。

智力这个变量则提出不同的测量可能性。这个变量的属性,并不仅限于对单纯分类的考虑:它所具有的属性暗含着相对大小的量。正是在这一方面,技术开始被采用。几个世纪以来,人们一直对如何测量智力这个问题感到困惑。相关努力包括,测试人们遇雨时是否懂得躲避——在这种情况下,智力可以被分成两类加以测量:那些懂得遇雨时躲避的人和那些不懂的人。然而,研究还在继续,而我们已拥有了智商测试。智商测试使我们了解到,人们如何能够很好地回答一些特定种类的问题,这些问题被认为与智力有某种关联。技术上的这种进步,使我们可根据与智力有关的量表,在其两个端点之间进行一定程度的细分。

测量本身自成一个研究领域。研究者一直想方设法发展测量技术,这些技术可以探究重要变量的所有属性。为了使我们对各种可能测量的理解系统化,科学家已经提出一种分类,这种分类包括四种测量层次:

1. 名目(定类测量)。
2. 顺序(定序测量)。

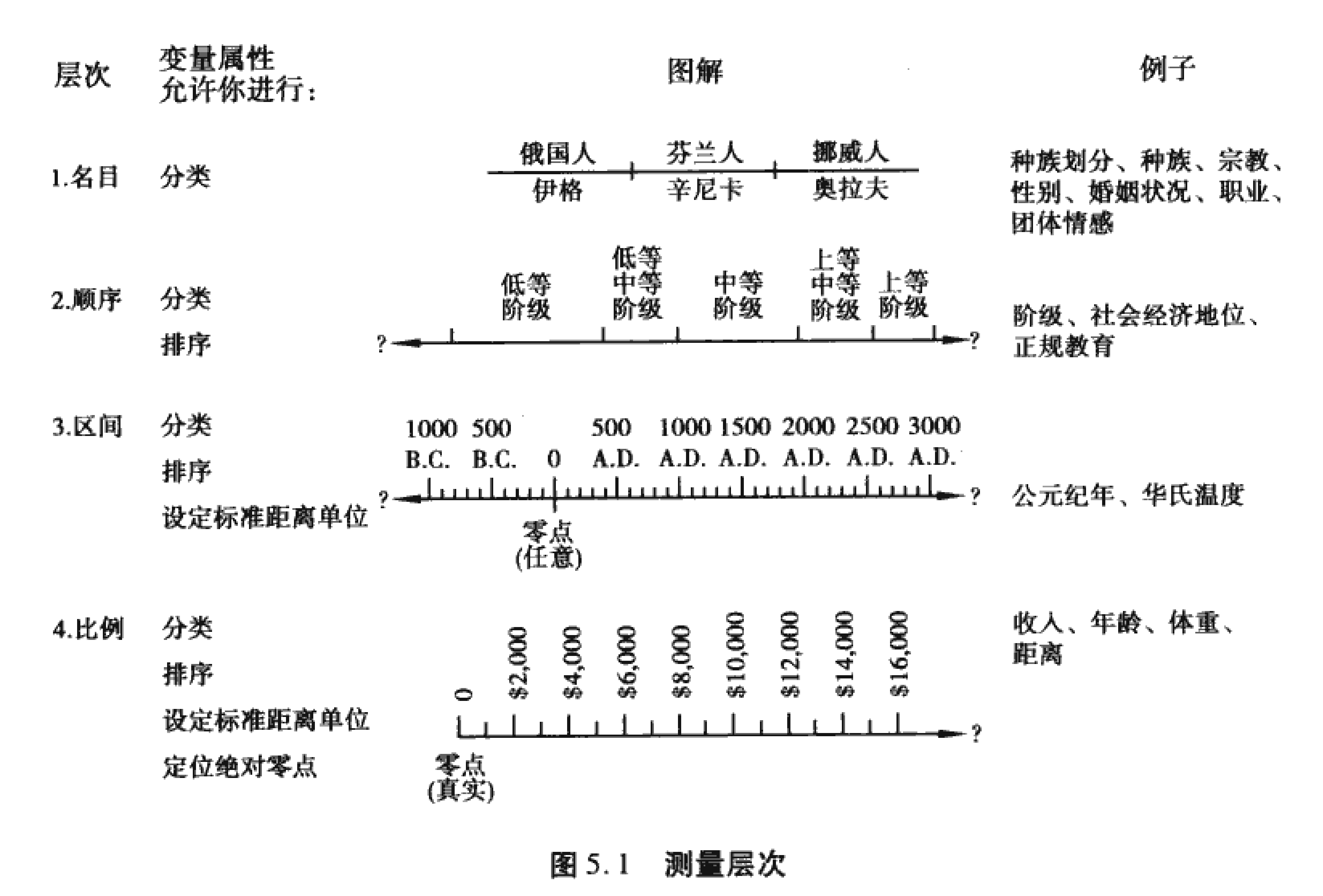


图5.1 测量层次

3. 区间(定距测量)。

4. 比例(定比测量)。

上述测量层次的特征在图5.1中得到明确说明。名目层次似乎与测量并不完全一样,它指的是事物的分类。以种族划分为例。假设辛尼卡是芬兰人,伊格是俄罗斯人,那么,我们说的就是两个人所具有的某种属性,这一属性与被称为种族划分的那个变量有关。这是测量,但并非十分花哨的测量。我们不可能把芬兰人排在俄罗斯人之上(除非根据另外某个变量——比如对腌制鲱鱼的喜好,即使是那样,他们的喜好也会很接近)。因此,作为一个变量,种族划分的属性只允许分类或名目测量。[①] 名目测量,尽管层次较低,但总是时不时地出现在社会科学中,如图中所列举的例子:种族、宗教、性别、职业,等等。

如果变量的属性除了允许分类之外,还允许排序,那么,只要有可资利用的技术,就可以尝试顺序层次的测量。在这一层次上,我们可以用某个连续统来思考,所谓的连续统就是一种排列,它能够显示变量的变化,与单纯分类不同。阶级是一个例子,社会经济地位则是另外一个例子。我们可以说,艾冯斯属于上层阶级,而迈克属于下层阶级。这些是分类,但它们按照从低到高的顺序被排列在一个连续统中。正式教育也一样:安琪莉娜拥有博士学位,玛丽受

① 在每一个伦理群体内,我们可以以门第为基础,鉴定某个人的遗产属于某一族群的比例是多少,但种族划分这个概念本身是分类性的。

过高中教育,而珍妮则拥有小学毕业证。但是,博士和大学学位之间的“距离”与高中文凭和小学毕业证之间的“距离”并不相同。这是排序,但并非标准距离。对距离的具体说明——或者更一般地说,事件之间的变化量——是迈向测量领域的重要一步。距离明显提高了测量的精密性,根据这种精密性,某一变量得以测量,并与其他变量发生关联。

倘若能落实标准距离,下一层次的测量便“应运而生”:**区间测量**。在此,我们可以指定测量单位,藉以说明各个事例相互间的距离有多远。这样做合情合理,但仍有一个技术细节引起我们内心的困惑。它与测量量表上的绝对零点有关。

区间测量没有真实零点。什么是真实零点?它有何用?以公元纪年为例,“零”年并非意味着,在那之前什么都未发生。我们确实不知道真实零点处于历史上的何处。零年是由于宗教原因在与基督生命的关系中被确立的,并且作为一个约定的参照系,起到前推后算的计年作用。华氏温度同样如此。我们都知道,华氏零度并不代表一个真实零点,因为零下23度比0度冷得多。

比例量表的确有真实零点。比例量表(比如距离)不同于区间量表,原因在于:例如,在一个比例尺上,零英寸就是表示零英寸——根本没任何距离可言。不可能有小于0的距离,或小于0的重量,或少于0的香蕉。上述这些向我们表明,真实零点与任意零点(或出于方便考虑被虚构的一个零点)之间的常规差异。

但是,真实零点有何用?答案取决于我们在比较比例或区间量表的观察时能够说些什么。如果哈蒂重200磅,劳瑞重100磅(比例量表),那么我们可以看出,哈蒂的体重是劳瑞的两倍。但是,假如星期一的气温是50度,星期三的气温是25度(区间量表),我们真的可以说,星期一的气温是星期三的两倍热吗?你可以试图这么说,并侥幸事实确实如此,但实际上你不应该这么说,因为上述那种比较需要一个*真实零点*。当你要辨别某天的气温是另一天的两倍热时,你需要知道何谓热的完全消失。没有一个起点,*距离可以被确立,比例则不然*。

了解这些区分的原因事关关系的种类,这些关系可通过统计方法在变量内及在变量间加以确立。任务就是要避免苹果和橘子之间不同类别的比较。本书所谈及的统计学,只以数字背后的观念的形式出现——统计演算以及各种统计操作的精妙之处,都留给更专门的论著。在此,我们只满足于简单的论点(就统计学的一般情况而言是很简单的)。大致说来:

> 在名目测量中,统计学处理的是每一类别中的事例频率(例如,种族划分:10个芬兰人,3个俄罗斯人)。
>
> 在顺序测量中,统计学描述事例依照某一变量的排序方式(例如,教

育：小学、中学、大学）。

在区间测量中，要在一个量表上对事例之间的量化差异进行比较（例如，年代：1950 年、1990 年）。

在比例测量中，要对事例之间的绝对距离进行比较（例如，钱：10 美元，20 美元）。[②]

由于这些测量层次对于处理变量关系显得至关重要，因此，在继续研究之前，有必要弄清楚每一变量的合适的测量层次。当我们转向这个问题时：测量以相关（correlation）的形式体现出来的变量关系，我们就会明白更详细地阐明测量层次的重要性。

测量数据的显著性和代表性：概率、抽样和民意调查问题

我们现在要讨论的这些主题之所以结合在一起，并不完全是因为它们都与测量有普遍的联系，而是因为，它们都与分析数据的长处和缺点相关。这些主题就是概率、抽样和民意调查问题。民意调查为研究抽样和概率提供了一个有益的平台，但这些主题在社会科学中还有更广泛的应用。

为了掌握科学研究必需的统计工具，我们有必要熟悉一个新概念：概率。概率在社会科学中占据的重要地位，远非本书几处描述就能表明的。概率构成了科学观点的基础。明了其原因，便能控制人类心灵中一些特殊的执拗习惯。

概率指的是某事发生的可能性或几率。我们计算下列事情发生的概率：通过一门课程考试的几率、对一场约会的展望、一个队赢得一场比赛的胜算。《罗热的同义词辞典》（*Roget's Thesaurus*）列举了足够多的表示概率的替代词汇——幸运、冒险、机缘、命运、偶然、几率，以及其他一些词汇，这一事实表明概率这个概念在我们语言中的重要性。

我们开始就说，科学作为一种手段，有助于人类应对生命的不确定性。将

---

② 摘自 Sidney Siegel, *Nonparametric Statistics for the Social Science*（New York: McGraw-Hill, 1956）, p. 30。另请参阅 Chava Frankfort-Nachmias and David Nachmias, *Research Methods in the Social Science*, 6th ed.（New York: St. Martin's: 2000）, Chapter 7。

对熟悉统计学的人而言，以下这些例子是适合于每一层次的统计量清单：

名目：样式、频数、列联表

顺序：中位数、百分位数、斯皮尔曼相关系数、肯德尔 T 系数、古德曼-库鲁斯卡的伽玛

区间：平均数、标准偏差、皮尔逊积差相关、多重相关

比例：几何平均数、变异系数、OLS 回归

思想观念从头脑中“驱逐”出来，迫使它们进入经验观察的领域，并且在经验中检验它们，我们就获得关于世界的知识。我们努力摆脱环境的不确定性带来的不安全感，正是这种努力，提供了科学体制赖以建立的动力。但是，科学知识的特征就在于，它很少会坚如磐石。时而明显时而隐晦，科学概推总是概率性的。

科学是几率的提炼，这一点要远比发现确定性更常见。事实上，社会科学家通常以一种语言来讨论他们的发现，这种语言表达了结果出错的可能性。他们为下列情况发生的可能性感到担忧：一组结果可能反映的是一个不精确的样本，或者，另外一位研究者可能会从我们的结果中发现不同的结果。规范地说，我们很少会认为，社会科学决定性地证明了某种东西。[3] 相反，我们经常会说一个假说得到有效证据支持的概率有多大。

概率以何种方式植根于社会科学之中？作为对这种方式的一种说明，我们将考察概率统计的两种具体应用：其一就是确定数据组的统计显著性；其二就是构建关于较大母体的代表性样本。但在这两种应用中，目标都是一样的：尽力阐明一列数据揭示的不仅仅是变量间随机关系的可能性有多大。如果信息建立在一个劣质样本的基础之上，或者，如果它仅代表事例的反常结合，那我们就不能说，结果向我们表明了变量之间有任何确定的关系。了解这一点是很重要的，而概率统计提供了一些工具。

概率的第一种应用关系到从一个较大母体中抽取出的某个样本的代表性。假若一个样本的容量和特征已定，那么，我们从这个样本可以推断某一母体的特定特征的概率有多大？这种类型的概率，构成民意调查的基础。民意调查人总是试图估计，打算投票给特定候选人的公众占多大比例。在尝试概括一个巨大人群的特征时，我们几乎不可能总是调查每一个人。选择最小的、最具代表性的可能样本，是有效进行民意调查的关键所在。概率统计被用来估计一个样本具有代表性的几率。

概率的第二种应用关系到估计一组观察随机发生的可能性。如果我们正在观察的结果，每 100 次中只有 1 次会随机发生，那么，这个模型就相当有意义。连接两个变量（比如，收入和教育）的某个数据模型，每 100 次中只有 1 次随机发生的几率，那么，这个模型告诉我们一些有用的信息。无需探究数学运算，我们便会认为，这一结果的显著性水平为 0.01。将样本的观察次数、变量的变化量，以及所观察的关系强度结合起来，就能推出显著性的统计量。最可能的随机结果分布将表明，一个表格的每一格中的事例数目都是一样的；最不可能的随机分布，则是所有事例都出现在一个格子中。

---

③ 来自科学哲学领域、对实证主义的一种有影响的批评，可参见 W. V. O. Quine, “The Two Dogmas of Empiricism,” in *From a Logical Point of View*, ed. W. V. O. Quine (New York: Harper & Row, 1961)。另请参阅 Alexander Rosenberg, *The Structure of Biological Science* (Cambridge, England: Cambridge University Press, 1985)。

确定结果的显著性水平，构成一次重要的假说检验。最不可能随机发生的结果是，所有高收入的人都受过高程度的教育，而所有低收入的人都受过低程度的教育。自变量对因变量有相当强的影响。在母体中，收入和教育之间都极有可能存在某种相关性。显著性检验告诉我们，在特定条件下，我们的假说是对（或错）的概率有多大。

就其最基本的形式而言，显著性告诉我们"某种特定的关系……是否值得深入思考——它是否值得我们付出额外的努力"。[4] 有些社会科学家可能只处理显著性水平为0.01的数据，而其他一些社会科学家则将0.05视为显著性水平的判定值——意思是说，观察到的关系，每100次中有5次随机发生的几率（与1次几率不同）。研究者已普遍注意到，显著性水平是研究报告的一部分，这有助于对结果进行评价。

我们一直在讨论的两种概率应用是相关的：前者处理的问题是样本是否具有代表性，后者则与结果具有意义的几率相关。笼统说，这个问题就是：

> 样本具有代表性吗？［推断］
> 结果模式可能会随机发生吗？［显著性］

一个具有代表性的样本，为推断某一假说的支持度提供了坚实基础，尤其是当观察模型具有统计显著性的时候。但是，一个糟糕的样本导致糟糕的推断，无论数据模型是否具有显著性。毫无疑问，根据概率论抽取的样本被认为是概率样本。

在抽样时所运用的技术，一般有两种：分层化和随机抽样。分层化涉及按比例展示样本中的重要特征，藉此再造一个大的母体。假设我们试图测定某一社区对饮酒的态度，在当地一家酒吧选择一部分顾客作为样本进行访谈，由于此样本的特征对所思考的议题显得至关重要，因此会使它所代表的那部分公众超过比例，因为禁酒主义者不会在酒吧泡着。这样一来，我们必须以另一种方式来选择样本：使得禁酒主义者亦有机会被包括在内。

如果运用分层抽样方法选取一份样本，以检验选举中的投票行为，我们会设法获得这样一份样本：它按比例反映出较大母体的特征——至少包括诸如阶级、宗教和教育等重要的自变量。但是，分层化（对投票者的特定特征进行比例抽样）必须被限定在一些数量相对较少的特征上。如若不然，为了按恰当比例填满所有变量的代表样本，我们可能会因为试图寻找那些极不可能集各种特征于一身的人们而耗费宝贵资源。

随机抽样有赖于随机选取一个充分的关于母体的样本，以便这个样本有较高的概率再现整个母体的本质特征。随着样本容量的增长，样本具有代表性的可能性也在增加。例如，假如我们从一个全国人口为2 100 000 000的国

---

④　关于显著性测试的一个明晰而又通俗易懂的讨论，请参见Lawrence Mohr, *Understanding Significance Tests*, Sage University Paper Series on Quantitative Applications in the Social Science, no. 73 (Beverly Hills, Calif.: Sage Publications, 1990)。

家中，随机挑选5个人进行访谈，那么他们并不真正具有代表性的可能性很高，也就是说，挑选时可能存在相当大的误差幅度。随着样本容量的逐渐增长，只要受访者是随机挑选的，误差幅度便会降低。

对任何大小的母体而言，都有可能在数学上确定一个既定的样本容量将有多大的概率，产生某种可具体指明的误差幅度。误差幅度随着样本容量的增长而急剧下降，达到某一点时，样本容量的进一步增长将非常轻微地减少误差幅度。正是这个点指示了最经济的样本规模。越过这个点，再增加两倍或三倍、甚或十倍的样本容量，所能降低的误差幅度还是会相当少。

随机抽样的一个主要问题在于，为了访谈所有被选中的人士，访谈者不得不分散他们的努力，并且不得不在整个母体的各个角落中寻找受访者。大多数科学抽样均同时运用分层抽样和随机抽样。例如，在一个全国样本中，研究者可以挑选有代表性的市区，以及有代表性的乡村（分层化的一种形式），然后在这些目标区域中抽取一个随机样本。

电话抽样，尽管对那些没有电话的人存有偏见，但已经成为一种日渐流行的技术，因为计算机使得在特定电话交换机内进行随机拨号成为可能。然而，通讯技术中的革命，正给电话抽样带来新的偏差。寻呼机、BP机以及电话答录机，都与大多数的电话交换机相连。公众对电话（电信）骚扰的恼怒可能会提高对调查研究者的拒访率。良好的电话样本仍可以被抽取，但随着研究者努力降低这些新的偏差的同时，成本也在提高。互联网为成立应答者会谈小组创造新的可能，我们可以就一些重要的问题对他们进行调查。

为了进行美国民意调查，一些主要的学术性、商业性，以及媒体民意调查（皮尤研究、盖洛普、哈里斯、有线电视新闻网/今日美国、哥伦比亚广播公司/纽约时报、ABC新闻/华盛顿邮报、美国全国广播公司/华尔街日报），均运用一种分层的随机样本，目的首先在于消除不方便的问题：比如，访谈一位远在内华达州的偏远地区随机挑选出来的牧羊人。样本容量通常约为1 500人。在这种大小的容量上，误差幅度在0.05显著性水平上约为3%。意思是说，就意见测量而言，95%的样本，在正负3%的范围内，可反映出实际母体。因此，假设49%的实际人口打算投票给民主党候选人，那么，95%的样本将产生出在46%～52%这个范围波动的估计值。⑤

运用这种类型的样本预测总统选举时，一些主要的媒体公司一直有着相当辉煌的成就纪录，部分原因在于，他们一直很幸运，未抽到一个“异常”样本（5%中的其中一个），还有一部分原因在于，他们的确对其样本进行了分层，一定程度上避免了异常样本的出现。倘若仅利用简单随机抽样，异常样本就可能出现。这并不意味着，他们的预测就相当准确。1996年选举之前，哥伦比亚广播公司/《纽约时报》立即进行民意调查，结果过高估计了克林顿获胜的票额差数，

⑤ 我们感谢 Albert Klumpp 提供这个例子。

高了10个百分点。《时代周刊》断定，它的"可能选民"样本是不恰当的。[⑥]

如果你试图利用某一样本来代表相互之间几乎无法区分的现实值，那么预测同样是困难的。2000年选举夜，一些主要的广播联通公司所运用的投票后民意调查的数据，导致他们对佛罗里达州的结果做出错误预测。佛罗里达州的一个问题在于，那里的投票结果相当接近——基本上是49%对49%。[⑦] 由于这些广播联通公司呼吁佛罗里达州支持戈尔，所以他们的民意调查数据必定表明，戈尔至少领先1到2个百分点，因此可以肯定，他们的预测超出了误差幅度。[⑧]

2000年佛罗里达州投票后，民意调查过分自信地表示，戈尔比布什赢得更多选票，这一点导致他的失败。这并不意味着，有关谁会赢得选举的预测就是完全错误的。投票后民意调查测量的是，投票者在选举日想将选票投给谁。相对多数的佛罗里达州的投票者，可能的确想在选举日投票支持戈尔——但是，投票者的过失、设备故障，以及选举结束后非法地将海外选票计算在内，这些都搅乱了最终的计数，并且可能剥夺了戈尔当总统的机会。[⑨] 即使将所有这些因素都考虑在内，这些民意调查还是过高估计了那些自信已投票给戈尔的选民的比例。不管怎么说，戈尔和布什之间的（所获选票）真实差额就不到0.05个百分点。一个极具代表性的样本应该已经表明了这一点，并且致使广播联通公司认识到，两人的选票是如此接近以至于无法在广播中播出（一个明确的当选结果）。

理解抽样的另外一种方式，就是思考劣质样本如何导致错误的结论。劣质样本在预测总统选举时曾导致一些严重的错误。一个现已停办的刊物《文摘》，由于每次选举都要对无数投票者进行调查，而且（精确地）预测了1924、1928和1932年的总统选举，所以一度闻名遐迩。许多人将这些预测的精确

---

⑥ 参见 Michael R. Kagay, "Experts See a Need for Redefining Election Polls", *New York Times*(December 15, 1996), p. A18。

⑦ 在美国最高法院介入停止重新计票以后，最终的官方合计数字表明，在佛罗里达州5 963 000选民中，乔治·布什的得票高于戈尔537张，即赢得48.85%的支持率。根据这个合计数字，戈尔赢得48.84%的支持率。布什在佛罗里达州赢得了相对多数的选举团的选票。

⑧ 投票后民意调查，通常会运用大的分层的因州而异的样本，这些样本降低了一些主要的媒体民意调查所使用的误差幅度的标准，即在上下3个百分点之间波动。

⑨ 对未被计算在内的穿孔卡片、"悬挂式选票"、"毛虫式选票"、"蝶形选票"的选举后的审查，以及被非法计算在内的"海外"选票，均表明：在佛罗里达州的选举中，想投票支持戈尔的人，要比投票支持布什的人多几千人。参见 Jonathan Wand et al, "The Butterfly Did It: The Aberrant Vote for Buchanan in Palm Beach Country, Florida", *American Political Science Review*, 95, no. 4 (2001): pp. 793-810. David Barstow & Don Van Natta Jr "How Bush Stole Florida: Mining the Overseas Ballots" *New York Times*(July 15, 2001). Kosuke Imai & Gary King, "Did Illegally Counted Overseas Ballots Decide the 2000 U. S. Presidential Election?" Manuscript. Department of Government, Harvard University. 2001。

表 5.1　税收和支出:样本偏差的影响

问题 1:你是否相信,税收每增加 1 美元,就应当出于赤字和债务削减的需要而节约两美元的支出?			
	是(%)	否(%)	没有答案(%)
《电视指南》回邮的回答	97	无资料	无资料
杨克维奇全国样本	67	18	15
问题 2:是否应通过法律,以排除特定利益集团捐巨资给候选人的所有可能性?			
	是(%)	否(%)	没有答案(%)
《电视指南》回邮的回答	99	无资料	无资料
杨克维奇全国样本	80	17	3

注:"佩罗(无意)策动了一项民意调查实验,"引自 David M. Wilbur《公共视点》(Vol. 4, #4©1993)第 28-29 页。再版经过了公共意见研究中心的许可。

性归功于庞大的样本容量。《文摘》从拥有小轿车和电话的人的列表中抽取一部分人,并用邮寄的方式对他们进行调查,然而,这种方法在 20 世纪 20 年代和 30 年代要冒很大的风险:它使得不怎么富裕的人的比例过低。1936 年,在对超过二百万的投票者进行调查之后,《文摘》根据其样本推断,阿尔夫·兰登(Alf Landon)将击败总统罗斯福,因为兰登将获得 57% 的选票。总统是兰登? 事实并非如此。兰登在选举中仅获得 36% 的选票。应当强调,一个大的样本并不必然抵消无代表性。[10]

类似于分层随机抽样的技术已大大提高了民意调查的可信度;但是,候选人和利益集团所进行的调查,有时还会重蹈《文摘》的覆辙。请看表 5.1 所示的数据。在 1992 年总统竞选的过程中,独立候选人罗斯·佩罗(Ross Perot)的集团,通过《电视指南》杂志将民意调查报告邮寄出去,并对读者做如下要求:当他们观看某一候选人的电视演讲时,就在他们的选票上做出相应的标记。随后,读者的反馈意见被寄回佩罗集团。然而,当另外一个集团采取随机概率样本,并向应答者提出同样的要求时,他们得到的结果与佩罗集团报告的结果大相径庭。[11] 佩罗集团的数据犯了样本偏差的错误;他们并未真正尝试随机化或分层化。因此,这种民意调查不可能成为民意的可靠代表。

---

⑩ 1936 年选举成就了乔治·盖洛普(George Gallup)。他用一个相当小的、相对随机的"定额"样本,预测罗斯福将获胜。至于对这次选举中的民意调查的说明,请参阅 David W. Moore, *The Superpollsters* (New York: Four Walls Eight Windows, 1992)。但是盖洛普 1936 年所使用的定额方式,却导致了另外一次有名的(预测)失败:他的样本导致他预测杜威将在 1948 年选举中击败杜鲁门。自 1948 年以后,盖洛普和其他的民意调查人最终采用了概率样本。

⑪ Daniel Goleman, "Pollsters Enlist Psychologists in Quest for Unbiased Results," *New York Times* (September 7, 1993), pp. B5, B8. 另请参阅 *The Public Perspective* (Roper Center for Public Research), May/June 1993。

表5.2　常见的误差来源

误　差	例　子
模糊问题:	你是否认为,我们应当为了和平或强大的国防而奋斗?
引起偏差答案的带有象征意味的问题:	你是否认为,都市儿童有生命权?你是否认为,怀孕妇女应当有权选择堕胎?
超出回答者信息范围的困难问题:	你是否赞同美国在与俄罗斯进行裁减军备的谈判中对于弹道飞弹部队所持的立场?
不适合问题主题的回答选项:	你感觉未来会更好还是会更差?
包含不止一项议题的问题:	你是否更有可能支持一位赞同校车接送以及强大国防的候选人,或者更有可能支持一位个性讨喜的候选人?

应当注意的是,除了样本的代表性之外,其他误差来源也悄悄地在调查研究中蔓延。某一研究者可能已经挑选了一个具有高度代表性的样本,但其测量工具可能导致误导性的答案。常见的误差来源包括表5.2中列举的那些。

除了这些明显的误差类型外,还有一些误差源自一种困境:你很难确定你在测量的东西就是你认为你正在测量的东西。举个例子,我们想了解人们对穷人的个人同情心,出于对此的兴趣,我们提出这样一个问题:你是否赞同穷人在饥饿时偷面包的行为?对穷人抱有极强同情心的某个人,可能会说"我不赞成",因为那个人在对穷人抱有极强同情心的同时,还极其尊重法律和秩序。请注意,这个问题并非没有意义,误差来源于把某种不恰当的意义赋予回答。除了预想的一个变量(对穷人的态度)之外,这个问题还提出了另外一个供讨论的变量——对法律和秩序的尊重。[12]

除了草率的测量所导致的误差之外,还有就是用来概括数据特征的统计程序所引起的误差。至少在一种较低的程度上,统计学往往会歪曲现实——这也正是统计学家为什么更愿意运用多种技术概括数据特征的原因所在,目的就是要避免单一程序所造成的偏差。

测量变量间的关系:关联和相关

关　联

确立两个或多个变量间的关联度,成为科学事业的核心目标。科学家花大量时间以阐明一个事物如何与另一事物相关,并将这些关系建构为说明性

[12] 对民意调查的一个通俗易懂的介绍,可参见 Herbert Asher, *Polling and the Public: What Every Citizen Should Know* (Washington, D.C.: CQ Press, 1998)。

的理论。

与其他形式的测量一样，关联问题也经常出现在一般的谈话中，比如在下面这些话中："有其父必有其子"；"窥一斑而知全豹"；"一天一个柑橘，远离坏血病"。在从统计上测量变量间的关联程度时，科学家只是做名副其实的事情：将司空见惯的行为严格化和精确化。

有时可以用一些简单方式来概括关联的特征。一个变量对另一变量的影响可以用文字或统计数字来描述。"用 Crust 牙膏的人，有更少的龋齿"这个陈述，表示了自变量（*用 Crust 牙膏刷牙*）和因变量（*龋齿的数量*）之间的关系。

在对关联进行具体说明时，可以有效地使用诸如中位数、平均数及标准偏差这类描述性统计量。例如，我们在第 2 章关于选举制度和投票人数的讨论中，每一国家中的选民参与的百分比大约可均分给三种不同类别的选举制度（见表 2.2）。这使我们能分析选举制度对投票人数的影响，而且，百分比的差异还是唾手可得的比较手段。例如，我们在第 3 章关于社团成员身份和世代之间关系的讨论（见表 3.3）表明，在最年轻的年龄群中，只有 22% 的人是三个或更多团体的会员。倘若年龄段和团体成员身份之间不存在任何关系，那么我们可能会以为，在每一年龄段中，都仅有 22% 的人加入三个或更多的团体。事实并非如此。在老龄群中，31% 的人加入三个或更多团体——二者之间差了 9 个百分点。

在这一点上，我们可能会问，9% 是否是一种足够大的差异，以至于我们可以说，年龄段和团体成员身份之间存在某种关系。在这种情况下，答案是肯定的。关于关联和相关的测量，为我们提供了回答这些问题的工具。[13]

关联和相关的测量

为了特定的应用，统计学家已经发展出更为精致的工具，以阐明变量间的关系：关联和相关的测量。不同层次的变量测量，需使用不同的统计方法来对关联进行测试。其结果就是一碗关于测试的字母汤（alphabet soup），它通常由希腊字母表中的字母标示，例如，*chi* 和 *cho*。所有这些测试总是共有某种通用的逻辑。

关联和相关的测量，通常被作为一种统计问题加以研究，在此，我们将集中讨论它们背后所隐藏的观念。当你看到某一相关统计量时，我们的讨论应该能帮助你认识到它。要理解算法和限制性假定，你可以参考统计学教材。

相关的基本观念，就是在统计上描述变量间的关联。假定所有其他条件都一样，那么，关联测量就可以概括彼此相关的两个变量的动向。

---

⑬ 在表 3.3 中，假定一个样本的容量为 14 259，样本中数据的卡方值为 150.1，那么这个样本将表明，群体会员数不依赖年龄群这种情况的发生概率相当低。

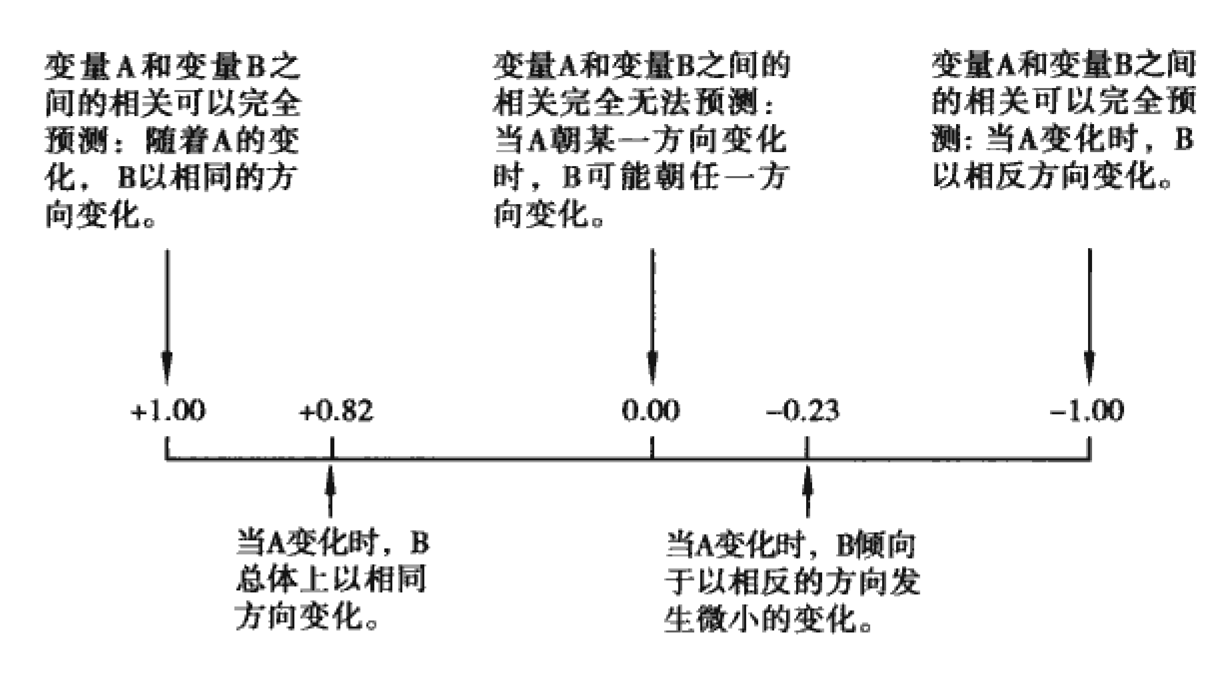


图 5.2 相关量表

相关分析作为一种进步，突破了百分比差异的比较，因为它允许你根据单一统计量来了解关联的方向和数量。方向指的是关联是正的还是负的。随着变量 A 的增加，变量 B 也增加，这时，二个变量间就存在一种正相关。也就是说，变量 A 随着变量 B 的上升而上升（反之亦然）。随着变量 A 的变化，变量 B 朝相反方向变化，这时，两个变量间就存在一种负相关。在负相关的情况下，随着 A 的增加，B 减少（或者，随着 A 的减少，B 增加）。

举个例子，一个气球中氦分子的数量与气球上升速度之间就存在一种正相关。上升速度和附在气球上的重量之间则存在一种负相关。正/负相关，用相关数字前的"+"或"-"表示。

相关强度用数值大小表示，这些数值的取值范围是从 0 到 +1.00 或从 0 到 -1.00。相关量表，可参见图 5.2 的说明。因此，相关统计提供了变量间关系，以及关联测量的简单指标——一种可能具有危险性的简单方法，因为这些统计数字背后的数学运算牵涉到需要谨慎思考的假定。除此之外，计算关联测量的各种技术，也会导致结果稍微偏离真实的数据。了解相关运算的一般技术，将使你能够看出一些（尽管不是全部）问题。尽管如此，在不完美的测量界，这些统计量仍然是有价值的工具。

用以计算关联测量的技术因所运用的不同测量层次而变化。如果在名目层次上测量两个变量（单纯的分类），那么，相比在区间层次上测量两个变量的情况而言，就很少可以对关联的特征进行概括。事实上，对每一种测量层次而言，都存在可供利用的相关技术，我们将对这些技术如何起作用进行一般的说明。

*名目层次上的关联。*名目测量只涉及简单的分类，是低等级的测量，而且与其相称的关联测量实际上不应当被称为相关。列联系数是一种统计量，通常用来概括数据的实际分布在多大程度上偏离两变量无关联时的分布。

假设一位研究者想检验一部古老的斯蒂夫·马丁（Steve Martin）的电影

片名所陈述的主张——死亡男人不穿彩格衣。为了透彻地检验，他决定检验死人和活人穿彩格衣的相对发生率，以观察死者是否确实有差异——就男式服装而言。这位研究者观看了二十份男性活人的样本，并在当地殡仪馆察看了二十具男性尸体。如果结果表明，死亡男人和活着的男人，穿彩格衣的情况一样常见，那么马丁的假设没太大价值。

表5.3呈现了一组结果，这些结果表明，穿彩格衣与男性的状况（无论是死者，还是活人）之间并没有关联。这正是我们可能看到的情况，倘若两个变量间没有关联的话。数据显示，24个穿彩格衣的男人，均分在活人和死人之间。同样，16位不穿彩格衣的男人，在活人和死人的分布也是一样多。在这种情况下，列联系数将是0。

但是，假设我们实际上观察到其他不一样的情况。假设死者和活人当中，穿着彩格衣的男性的数量存在差异，结果又会如何？如何用精确的统计术语来表达呢？

列联系数的计算是根据以下两种数据分布的比较得来的：(1)如果变量间不存在关联，我们预期观察到的值（表5.3）；(2)我们实际上观察到的值（表5.4）。表5.3的结果，可以表达为卡方（chi-square）统计量，或列联系数。由于卡方没有上限，所以出于解释的考虑，它通常被转换为类似于列联系数的统计量，较大的值反映较强的关联。如果我们发现，数据分布的确如表5.3所示，[14]那么，卡方值就等于8.12，列联系数将是0.41。这些统计量指出，死亡与穿着之间存在某种关联——但并非电影片名中所指出的那种关联。相比活着的男人而言，穿彩格衣的死亡男人的比例似乎更大。

表5.3　穿彩格衣的预期发生率

穿着类型	男性的状况	
	活着	死亡
彩格衣	12	12
非彩格衣	8	8

表5.4　穿彩格衣的实际观察到的发生率

穿着类型	男性的状况	
	活着	死亡
彩格衣	5	14
非彩格衣	15	6

[14] 参阅 G. W. Bohrnstedt & D. Knoke, *Statistics for Social Data Analysis* (Itasca, Ill.: F. E. Peacock, 1988), p. 310。

用更系统的语言来说，这种数据分布指出，死者和穿着特定格调的衣服（在此例中，就是彩格衣）之间存在某种关联。列联系数被用来概括非排序性的、名目层次的变量间的关联。[15] 这个测量值在 -1.0 到 +1.0 这个范围内变化。记住，负统计量表明，随着一个变量的增加，另一个变量便会减少（反之亦然）。正统计量意味着，两个变量朝着相同方向变化。对变量中的类目进行排序的同时又进行分类，当这种可能性存在时，确立真正的相关性就成为可能。

顺序层次的相关。“顺序”意味着排序。这一特征为顺序层次上计算的统计量提供了基础。我们可以根据事例依照两个变量的排列顺序来比较它们的等级。举个例子可能有助于说明这一点。

设想，160 名童子军组成的一个团体，高唱着“上帝保佑美国”。领唱者（一个马克思主义者）按照歌者的音乐才能，将他们从最好到最差的顺序划分为四类：金丝雀、知更鸟、麻雀和乌鸦。他打算检验他的信条：下层大众比贵族阶层唱得更好。

因此，他可同时使用两种顺序分类法：根据金丝雀、知更鸟、麻雀和乌鸦，对*音乐才能*进行排序；根据上层、中上层、中下层和下层，对*阶级*进行排序。他打算检验的假说就是：社会经济阶级与音乐才能之间是否存在某种关联。那个领唱者假定，下层阶级的人比上层阶级的人唱得更动听。

如果上述假设为真，那么，数据将表明一种特定模型。随着阶级的上升，音乐才能将下降。下层阶级的人，将集中分布在“金丝雀”这一类别，而上层阶级的人集中分布在“乌鸦”这一类别。假定他所发现的数据分布，呈现在表 5.5 中。

从这些数据看，这种关系不是十分清楚，但是我们可以看出，下层阶级的童子军有一种显著倾向：他们的歌声比“上等人”的歌声更悦耳。现在，我们需要一种有助于确定关联度的统计量。除了其他类似的统计量之外，古德曼-库鲁斯卡（Goodman Kruskal）的*伽玛*运用一种有意思的逻辑演算来概括关联度。*伽玛*反映的是：已知自变量（阶级）的等级分布，在预测因变量（音乐才能）的等级分布时误差减少的比例。如果某人的阶级地位不折不扣地预测了其歌唱才能，*伽玛*就会很高；若非如此，*伽玛*就会很低。

表 5.5 音乐才能与社会阶级

才能	上层阶级	中上层阶级	中下层阶级	下层阶级
金丝雀	0	0	5	30
知更鸟	0	10	20	10
麻雀	5	15	15	0
乌鸦	35	15	0	0

[15] 其他以卡方为基础、在 0 到 1 这一数值范围内、在名目层次上评估所测量的变量间关联的统计量是克拉马的 v 系数和 phi 系数。

回到这个假说：当我们提高阶级的等级时，数据是否表明，音乐才能的等级也相应的下降？在表 5.5 中所呈现的数据中，伽玛将是 -0.93。这意味着，如果我们知道一个人的阶级，它将使我们预测此人音乐才能等级的能力提高 93 个百分点（相比我对这个人的阶级地位一无所知的情况下的预测能力而言）。这一点是否证实这个假说？是的。阶级层次和歌唱才能之间存在一种负关联。

区间和比例层次的相关。为了进行区间或比例测量，你必须能够确立分析单位之间的距离。根据金丝雀、知更鸟、麻雀和乌鸦将歌手排列起来，这还不够好；金丝雀与知更鸟之间，以及与其他类别之间的距离量必须得到具体说明。金丝雀和知更鸟之间歌唱才能方面的差异，可能与麻雀和乌鸦之间的差异完全不同。随着距离的阐明，我们便有可能运用一种相关统计量，它可以利用距离因数来测量变量间的关联。

区间和比例测量允许我们使用一种出类拔萃的统计量，它被命名为皮尔逊积差相关系数，或皮尔逊相关系数。

为了化繁为简，我们将编造一个相当基本的例子：私有油井数量和卡迪拉克的数量之间的关系。我们的"样本"由 5 位拥有油井的业主组成。为了弄清楚皮尔逊积差相关的数学运算可实现什么，我们来考虑两组可能的数据。首先，假定油井数量和卡迪拉克数量之间存在 +1.00 的相关。我们可以宣称，图 5.3 所说明的两组数据，同样具有 +1.00 的相关。

请注意，图中所划的连接每一事例的实线表示两个变量间的如下关系：油井每增加一口，卡迪拉克便增加两辆。假如这条直线在一个不同的水平面上开始下落，便可得到完全相关的结果；例如，请看那条虚线。它表明，你在没有油井的情况下，也可得到一辆卡迪拉克，但是，每额外增加一辆卡迪拉克，似乎就有必要挖一口新的油井。

现在，我们设想这样一组数据，在其中，事例并未出现在一条直线上。倘若数据分布确如表 5.6 和图 5.4 所示，我们便无法划出任何一条直线来连接所有事例。假设存在一条直线，它最接近图中各点，亦即，它将所有事例偏离这条线的距离最大限度地缩小。通过一种数学处理，皮尔逊相关系数可以认定，一条表示线性关系的想象直线周围，点群（points cluster）靠近该线的紧密程度。由于数学上的原因（最好留给数学家），事例与这条线之间的偏差，根据距离的平方（$a^2+b^2+c^2+d^2$）来计算，而非单纯的距离之和（$a+b+c+d$）。事例与最恰当的那条线之间的距离越远，变量 A 与变量 B 之间的相关度便越低。就图 5.4 来说，皮尔逊相关系数为 +0.85。

皮尔逊相关统计可以被用来提供另外一条重要信息。对皮尔逊相关系数进行平方，我们就可发现，自变量的变化所解释的因变量的所有变化的比例是

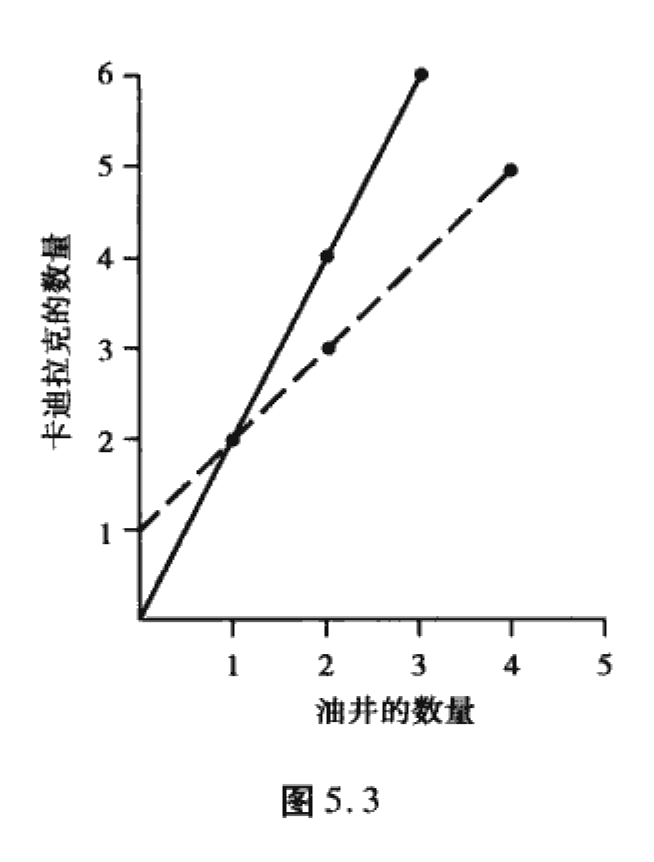


图5.3

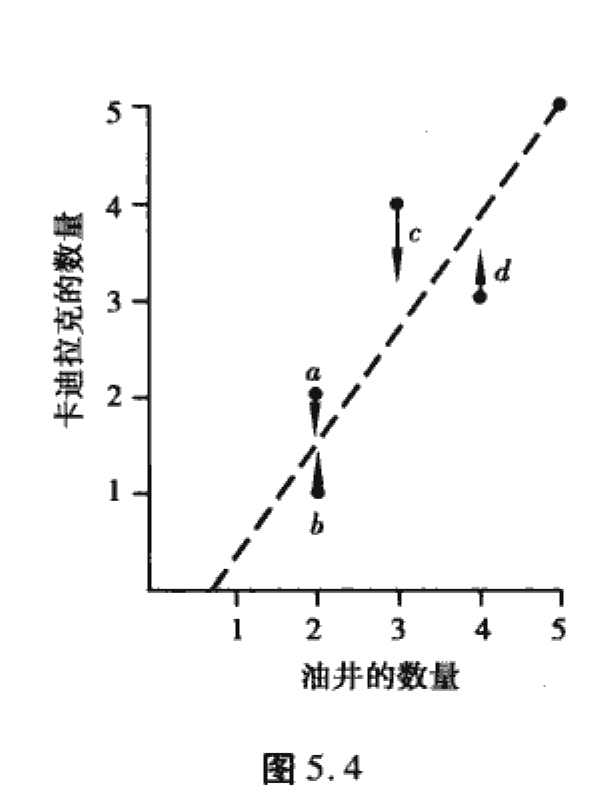


图5.4

多少。以图5.4中油井和卡迪拉克之间的关系为例，皮尔逊相关系数(r)是+0.85，所以r^2就是0.72(0.85×0.85)。因此，某人所拥有的油井数，可解释72%的私有卡迪拉克数量。其他变量则解释剩余28%的变化。

另外一种可能，假设所有点都落在这条线上(如图5.3所示)且$r = +1.00$，那么r^2还是等于1.00(也就是说，1×1=1)！油井数量与卡迪拉克的数量完全相关，并且没有任何变化需留待其他因素解释。

就此处所讨论的每一种关联测量而言，皮尔逊相关系数(r)告诉我们的是相关，它可能也可能不暗示某种因果关系。具有显著性的r并不表明，油井数导致人们拥有更多的卡迪拉克，仅仅表明，这两样东西是结合在一起的。实际关系可能是颠倒的(尽管在这个例子中，这一点似乎不合逻辑)。我们可能发现两个事物是相关的，但对哪一个变量决定另一个变量，并没有明确的想法。假设一位研究者发现，教育上的造诣与智力相关。那么，哪一个导致另一个呢？关联测量，并不需要研究者对因果关系做任何假定。统计学并不确立因果关系；因果关系取决于关系的逻辑(见表4.4)。

请注意，这套程序并不实现其他东西。皮尔逊相关系数为+1.00，这仅仅表明，A的任何变化都与B的一致变化有关。它并未向你表明，B随着A的变化而发生相关变化的单位数量是多少。在图5.3(实线)所示的例子中，所发生的情况的确是这样的：油井每增加一口，卡迪拉克便增加两辆。但是，假设情况是这样的：油井每增加一口，卡迪拉克就相应地增加1辆，或1/2辆，或3辆，这样一来，仍可得到+1.00的结果。

表 5.6 油井和卡迪拉克

卡迪拉克的数量	油井数				
	一口	二口	三口	四口	五口
一辆		1			
两辆		1			
三辆				1	
四辆			1		
五辆					1

用数学术语来说，皮尔逊相关系数仅告诉你一条想象直线周围的事例分布。它并未告诉你这条线的斜率，换言之，它并未告诉你，A 每变化一个单位，B 的变化量是多少。一种涉及先进统计概念的不同的统计程序，即所谓的*回归分析*，处理这个问题。

回归分析

这种普遍类型的测量方法旨在刻画变量间的相互影响。回归分析为这些特征增加一种新的精度。借助回归分析，如果你知道一个自变量的值，你就可以开始预测因变量的值。

有两种基本的回归类型：*双变量*回归和*多重变量*回归。与相关分析一样，双变量回归阐明，单一自变量的变化程度如何与一个因变量的变化相关。多重回归考察的是，若干不同的自变量如何与一个因变量发生关联。

作为双变量回归的一个例证，我们来思考职棒大联盟的一个简单例子。既然职业棒球队在球员工资方面开支不同，我们可能想了解，高额支出是否与球队的最终排名相关。球队薪金总额对胜率有多大影响？双变量回归根据下列线性方程式，概括出两个变量间的关联，可帮助解答上述问题。

$$Y[\text{胜率}] = a + bX[\text{球队薪金总额}]$$

用文字来表达这种关联，就是说：变量 Y（*球队胜率*）的值，是某一常量加上一定数量的变量 X 的一个函数。我们感兴趣的问题是，球队支付给队员的薪金总额是否影响到它在赛场上的成功。换言之，变量 Y（*胜率*）的变化，有多少与变量 X（*球队薪金总额*）一个单位的变化相关联。答案取决于 b，即回归系数。[16]

我们可以运用 2002 年棒球赛季的资料来检验这一假说：较高薪金总额的

---

[16] 用代数术语来说，b 代表 X 和 Y 之间关系的斜率。将代数与棒球结合起来，若我们将 Y 与 X 的关系绘制在图表上，那么 b 就是“纵距与横距之比”。

球队将赢得更多比赛。[17] 针对30支不同的大联盟球队，我们所测量的球队薪金总额（变量 X）是一支球队付给每位队员的平均工资，以百万美元为单位。我们所测量的胜率（变量 Y）是这支球队在2002赛季获胜场次的百分数。这两个变量间的皮尔逊相关系数是0.60，这表明，球队花费的金钱越多，它们赢得比赛的场次也就越多。双变量回归表明：

$$\text{球队获胜百分比} = 36 + 6.1 \times \text{球队薪金总额（以百万美元为单位）}$$

这意味着，一支球队的薪金总额和获胜百分比之间关系的斜率是6.1。这一结果的显著性水平是0.01。换句话说，平均每付给一位队员一百万美元，一支球队赢得比赛的场次就会增加6个百分点。一个常规的棒球赛季下来，这支球队将多赢得大约10场比赛。

这些结果表明，一支球队平均付给每位队员三百万美元，那么，我们就可以预测，这支球队将赢得54.3%的比赛（36 + 6.1 × 3 000 000美元 = 54.3%，或者一个0.543的纪录）。我们还可以表明，任何一支特定球队，如若想将其获胜百分比提高10个百分点（36 + 6.1 × 1 640 000美元 = 10.004%），就必须付给每位队员1 600 000多美元。[18]

以球队薪金总额为基础建立获胜百分比的模型，可解释整个30支大联盟球队获胜百分比中36%的变化。这意味着，其他变量——老练的经理、实力很强的投手、报酬不是很多的有天赋的年轻人，以及其他一些不太好捉摸的变量——在解释为什么一些球队获胜，而另外一些球队失败时，可能更有说服力。[19] 这个例子表明了我们有时候在社会科学中发现的一个要点：我们的回归模型尚有大量有待解释的地方，因为我们所模拟的一些东西，可能是一些无形因素的结果，这些无形因素很难被量化。

---

[17] 球队最终排名的数据来自 http://www.mlb.com。球队薪水的数据来自 http://www.sportingnews.com。

棒球作家保持着一种奇怪的传统：根据“百分率”来报道一个队的纪录，事实上，当他们这样做时，他们实际描述的是比例。也就是说，当一个队的胜率超过“0.500”时，这个队才算获胜。我们的数据作为真实的百分比被记录在案，所以获胜也就意味着纪录超过50.0%。如果我们将比例（而不是百分比）融入我们的分析中，那么 b 就要用比100次更小的单位来表示。

[18] Y 轴截距的常数（36%）代表的是：如果 X 是0，那么 Y 的值将是什么——在这个例子中，代表的是一支由优秀业余球员所组成的一支未付报酬的球队能够在一个赛季赢得比赛的百分比。

[19] 对这个例子而言，由于皮尔逊相关系数为0.60，所以我们知道，开支解释了球队胜率中36%的变化（0.60 × 0.60 = 0.36）。r^2 所代表的模型拟合度为什么正好是0.36，一个原因就在于，薪金总额低于平均数的大量球队在2002赛季都有最好的纪录（the Anaheim Angels and the Oakland A's），但是几支薪金总额较高的球队，其纪录却很糟糕（the Texas Rangers and the New York Mets）。

我们来思考另外一个例子。在附录 B 中，大卫·布罗金顿（David Brockington）、托德·多诺万（Todd Donovan）、肖恩·鲍勒（Shaun Bowler）、罗伯特·布里斯切托（Robert Brischetto）探究了下述问题：一种新型的地方选举制度如何能够导致种族性及民族性的少数族群的代表性。在标准的美国“赢者通吃”式的选举规则下，非洲裔美国人和拉美裔美国人通常很难有机会被选入市议会和学校董事会。即使是在少数族群占社区总人口的 30% 或 40% 的一些地方，一旦白人候选人获得相对多数票，并赢得所有在全市范围内竞选的“不分区”席位时，少数族群的代表也会被排除在外。在此，研究问题直截了当：在替代性的“累积投票”制下，非洲裔美国人和拉美裔美国人赢得的席位，是否与他们在社区总人口中所占的份额成正比？

由于只有少数美国社区直到 20 世纪 90 年代才采纳累积投票制，所以就这个问题先前几乎没什么研究。这项研究的发起人注意到，从理论上看，累积投票制理应产生少数族群代表，其比例大致与他们在某一城市的总人口中所占的份额成正比。[20] 先前研究已证明，比例性可以根据下列方式得到评估，即运用双变量回归来研究下述两个变量间的关系：其一就是社区中少数族群居民的百分比（自变量，或 X）；其二就是少数族群所占据的议会席位的百分比（因变量，或 Y）。

这项研究的发起人发现，对于那些少数族群是非洲裔美国人的地区而言，两变量间关系的斜率是 0.95，Y 轴的截距几乎等于 0（见附录 B 的表 2）。这意味着，在使用累积投票制的地方，非洲裔美国人的人口每增加 1 个百分点，他们所占据的议会席位的比例就相应地增加 0.95 个百分点。换言之，代表性接近于比例性。在附录 B 的图 1 和图 2 中，这种关系以图解的形式呈现出来。这些发起人将他们的结果与其他的关于少数族群代表性的回归分析的结果进行比较。通过比较，他们指出，在美国南部，相比于运用同一种族选区为基础的选举制度而言，这些新的累积投票规则甚至会产生更多的少数族群代表。

当回归牵涉到一个以上的变量时，将会如何（直接与政治相关的一个例子又会如何?）。举个多重回归的例子，假设你注意到，在选举政治方面，你朋友的参与度差异很大。活跃性可能涵盖此处举例说明的范围：

非投票者　　偶尔投票者　　政党活动家

为什么一些人比其他人更活跃？在行动主义之前（或独立于行动主义）所发生的因素，可能包括收入水平、教育程度或先前对政治的经验水平（称它们为自变量）。每一种因素在逻辑上都可能与积极性程度相关。多重回归可

---

[20] 在累积投票制下，投票者可能会投出与竞选席位一样多的选票。如果 4 个席位有待选举，那么某人可能会 4 次投票支持一位候选人，或者一次投票给 4 位候选人，等等。相关讨论，请参阅 Lani Guinier, *Lift Every Voice* (New York: Simon & Schuster, 1998)。

以同时分析若干自变量的影响。这种技术将单一自变量的影响分离出来，而同时又限定（或使其保持不变）其他自变量的影响。

为了继续考察这个例子，我们从一个人所参与的政治活动（包括投票、竞选、捐助候选人、参与抗议或示威，等等）的数量中得出一个分值，根据得分的多少，我们就可以对政治行动主义（因变量）进行具体操作。参与的活动越多，参与得分也就越高。简单起见，我们姑且假定，我们对政治行动主义的测量值，在0到10这个范围内变化。

如先前所见，相关测量使我们得以了解到，（比如说）收入和政治参与之间是正相关还是负相关（随着收入的增加，参与是上升还是下降），而且，相关还可以确立，一个变量如何紧随另一个变量而变化。

在对收入和参与之间的相关有所了解的情况下，你可以估计出，相比穷人而言，富人是否在政治上更积极。但是，假如你想拥有更高的几率来预测参与的变化量与收入水平的一个单位的变化之间的关联，就需要回归分析。在对一个或多个额外变量的影响进行限定的同时，若想检验一个自变量对一个因变量的影响，就需要多重回归。

例如，我们可能料想，政治行动主义的变化同时与收入及教育相关联。多重回归从统计上估计每个变量对因变量的单一影响而同时使其他自变量保持恒定。因此，在限定收入水平的同时，我们可以了解教育是否影响参与。多重回归表格通常会标明所分析的每个自变量的 b 值（斜率或回归系数）。它们可以被解释为：在其他被包括进来的变量（X_2，X_3 等）保持不变的情况下，与一个自变量（X_1）的一个单位的变化相关联的一个因变量（Y）的变化量。

回想一下，我们在第4章关于控制和虚假性的简短讨论。社会科学家经常无法运用实验室来控制或使多重变量的影响“保持恒定”。例如，将特定收入水平和教育程度随机分配给不同的人，然后将他们置于观察中，并等着观察参与是否以不同的比例出现在具有混合收入和教育水平的人们当中，这样做不仅不可能，而且也是残酷的。基于上述原因，社会科学家通常测量现存的现象，然后运用统计程序（比如多重回归）来控制那些无法（或不应该）被操纵的变量的影响。[21]

当自变量根据不同单位得以测量时，对多重回归系数的解释可能会变得很复杂。由于收入测量（美元）的范围远比教育测量（上学的年限）的范围要

---

[21] 处理回归的数学演算超出本书讨论的范围，但就这一主题，有无数的教材可供参考。简洁易懂的介绍，可参阅 L. Schroeder, D. Sjoquist, and P. Stephan, *Understanding Regression Analysis*, Sage University Paper Series on Quantitative Applications in the Social Science, no. 57 (Beverly Hills, Calif.: Sage Publications, 1986)；另参阅 M. Lewis-Beck, *Applied Regression*, Sage University Paper Series on Quantitative Applications in the Social Science, no. 22 (Beverly Hills, Calif.: Sage Publications, 1980)。

广，所以很难就这两个变量的回归系数进行比较。但是，一种统计技术可使它们具有可比性。比较单位可以用标准偏差单位表示。

"常规的"变量分布，大约 2/3 的观察值会落在中位数上下一个单位的标准偏差范围内。最远端的分值，位于中位数上下两个或三个单位的偏差范围内。

若一个变量有大量事例位于变化量的最高与最低端，那么，无论变化范围有多大，该变量都将具有较大的标准偏差。若一个变量的分值都集中在平均数周围，它将具有较低的偏差。这种将偏差标准化的方法，通常被用于回归分析。[22]

随之而来的标准回归系数一般指的就是贝塔系数；与皮尔逊相关系数一样，它通常在 -1.00 到 +1.00 这个范围内变化。就我们例子中的因变量（政治行动主义）而言，如果我们发现，收入的贝塔系数是 +1.00，那么这将表明，收入的一个标准偏差单位变化与参与的一个标准偏差单位变化相关。如果我们发现，贝塔系数是 +0.50，我们就可以推断，收入的一个标准偏差单位变化与参与的半个标准偏差单位的变化相关联。

回归系数和贝塔系数是回归分析的一个组成部分。另外一部分是多重相关统计量，即 R（不要同皮尔逊的小写的 r 混淆，小写 r 处理的是单一自变量的问题）。R 指明一组自变量和一个因变量之间的相关。与 r 一样，我们也可以对 R 进行平方，R^2 表示，由所思考的一组自变量所解释的一个因变量的变化比例。

总而言之，如果我们运用计算机统计软件包，就收入和教育对政治行动主义的影响进行多重回归分析，那么，我们的结果将包括每一个自变量的 b 值和 *beta* 值，而且还包括 R^2，它描述了所有自变量的影响。在我们的例子中，由于我们有两个自变量，所以我们就会有两个独特的 b 或斜率，且有两个独特的 *beta*。b 用什么单位表示，因变量就根据什么单位测量。因此，假设教育的 b 值为 0.25，这将意味着，在收入的影响保持不变的条件下，教育每发生一个单位的变化，政治行动主义的分值就将提高 1/4 点。假设教育的 *beta* 值为0.10，这将意味着，教育的标准偏差每提高一个单位，政治行动主义就会增加 0.10 个标准偏差单位。如果 $R^2 = 0.56$，我们便可得知，这两个变量所解释的政治行动主义测量的变化刚好超过一半。

我们对已经提出的这些关于测量变量间关系的统计量进行了一个归纳，并以此展开下面的议题，请参见表 5.7。

㉒ 运用标准回归系数带来大量技术上的问题。一些学者建议，研究发现应当只根据"真实"单位来表示。参阅 Gary King, "How Not to Lie with Statistics," *American Journal of Political Science* 30, no. 3 (1986): 666-687。

表5.7　关联、相关和回归的测量

统计量*	意　义
r	皮尔逊相关系数：两个变量间的共变度（-1.00～+1.00）。
r^2	决定系数：在双变量回归中，自变量的变化量所解释的因变量的变化比例（0～+1.00）。
R	多重回归系数：与两个或多个自变量的变化相关联的一个因变量的共变度（0～+1.00）。
R^2	决定系数：两个或多个自变量的变化所解释的一个因变量的变化比例（0～+1.00）。
b	非标准回归系数：自变量每发生一个单位的变化，一个因变量所发生的变化量（0～∞）。
$beta$	标准回归系数：一个或多个自变量每发生一个单位的变化，一个因变量的变化总量，其中，所有变量的单位都可以根据平均数之间的标准偏差进行比较（通常在-1.00～+1.00之间变化）。

* 出于本表需要，我们把皮尔逊相关系数作为唯一的相关测量来使用。至于其他测量，参见本章注②。

为什么要多重回归

控制和虚假性

在其他自变量的影响保持不变的同时，你如何确定某个自变量的独特影响？多重回归分析就是这个问题的答案。

请回顾第3章的讨论，研究社会资本的学者对以下这个问题感兴趣：自愿团体的会员身份，如何与政治参与相关联。在附录A中，罗伯特·普特南所提供的证据表明，低程度的团体会员身份与高程度的电视观看相关联。他写道："即使在对教育、收入、年龄、种族、居住地、工作状况和性别进行限定以后，电视观看与社会信任及团体会员身份之间还是具有很强的负相关关系。"

附录A的图X，对以下这一点提供了一种直观说明：社会科学家如何检验，一个变量对另一个变量的独特影响（在这个例子中，就是电视观看对团体参与的影响）在别的变量的影响被考虑进来时是否保持原样。根据普特南所提供的例子，我们可以发现，在各种教育程度上，看电视最少的人加入的团体也就最多。换言之，即使我们对教育的影响进行限定以后，电视观看和团体参与之间的负相关关系，还是保持不变。

为了解释美国社会资本的下降，普特南打算单独"揪出"一个变量，将其作为首要的嫌疑对象。在附录A中，大量证据表明，电视是"罪魁祸首"。通过对许多可能的竞争性说明的限定，他得以做出一种较强的断言：电视的有害影响不能归因为另外一种尚未说明的因素。这就是虚假性问题。

如果一项结果可被其他某个变量所解释,那它就被认为是虚假的。虚假关系的一个经典例子就是,火灾现场的消防车数量与火灾所导致的损失量之间的高度相关性。结论似乎很简单:增加消防车造成更多损失！被遗漏的一个变量毫无疑问是火势大小。较大的火势会引来更多的消防车,也会造成更大的损失。

回到社会资本的问题,倘若频繁看电视的人就是这样一些人,他们几乎没有任何时间或精力加入社会团体,那又会怎么样呢？情况可能是这样的:许多人现在要工作更长的时间,或者在往返途中要花去更多时间。当他们回家时,时间可能已经太晚了,以至于碰不到任何团体。倘若真是如此,他们最终也只好通过看电视的方式来打发夜晚时光。从这种观点看,电视观看并不是导致美国社会资本下降的首要原因。相反,它可能意味着另外一个重要变量(即缺少自由时间),而这个变量才导致了人们放弃加入社会团体。

任何科学事业的一个根本目标,就是对其他一些变量进行限定,这些变量可能会通过解释消除某一重要的结果。在你本人的研究中,你可能不运用先进的统计或实验室的实验限定某一个或所有被遗漏的变量,而这些变量也可解释你的观察结果。尽管如此,你还是应该考虑一下,你如何确信,你的结果不是虚假的。

表 5.8　法官种族对量刑严重性以及判处监禁的影响

	判处监禁		量刑严重性		
	MLE	MLE/SE	*b*	*beta*	*t*
所有被告					
没有限定	0.10	2.10*	-0.48	-0.01	-0.67
限定被告和罪行	0.14	2.17*	-0.91	-0.02	-1.60
限定法官、被告和罪行	0.11	1.67	-1.22	-0.03	-2.14*
白人被告					
没有限定	0.11	0.92	0.27	0.01	0.20
限定被告和罪行	0.35	2.19*	1.40	0.04	1.31
限定法官、被告和罪行	0.35	2.18*	1.39	0.04	1.30
黑人被告					
没有限定	0.10	1.72	-0.80	-0.02	-0.96
限定被告和罪行	0.09	1.24	-1.59	-0.04	-2.40*
限定法官、被告和罪行	0.06	0.78	-2.00	-0.05	-2.99*

编码:黑人法官 =1,白人法官 =0。对被告和罪行的限定包括罪行严重性、无论被告是否认罪悔改、被告的前科,以及被告是否有公设辩护人。对法官特征的额外限定包括:诉讼经验、性别,以及任法官的年限。对所有被告、所有白人被告以及所有黑人被告而言,*N* 分别等于 3 418,763,2 655。

* = 显著性水平为 0.05。

注:MLE 就是极大似然估计;SE 就是标准误差。

来源:Susan Welch, Michael Combs, and John Gruhl, "Do Black Judges Make a Difference?" *American Journal of Political Science* 32, no. 1 (1988): 126-136.

为了更深入地讨论,我们来看另外一个例子。表5.8取自威尔奇、坎布斯和古尔(Susan Welch, Michael Combs, John Gruhl)所进行的一项研究。他们探究了这个问题——"黑人法官是否有影响?"他们考察了影响主审法官量刑裁决的一些因素。这些研究者想知道,白人法官对待刑事被告的态度是否与黑人法官有所不同。这里的研究问题很明确:量刑裁决是否受影响于法官的种族?

先前研究不具有决定性——不同的研究发现了相互矛盾的结果。这项研究的发起者注意到,先前的研究者未能限定一些重要的自变量,比如,刑事被告的罪行严重性、被告的前科,以及法官的其他特征(如性别)。

威尔奇等人在东北部的一个大城市中,从1968—1979年间被判处犯有重罪的男性被告样本中发掘出他们的数据。在其中一部分分析中,他们将因变量操作化为法官量刑的严重性。他们使用了严重性量表,在其中,零相当于缓期执行;较低分数代表罚款和保护管束;较高分数代表监禁服刑时间;最高为93分,相当于终身监禁。因为他们想对多个自变量进行限定,所以他们运用多重回归分析。

运用某一类型的双变量回归,他们发现,法官的种族(自变量)与判决严重性并没有关联。也就是说,当他们仅仅着眼于法官的种族和量刑严重性之间的关系时,他们发现,二者之间并没有关联。[23] 但是,当他们运用多重回归并且对被告罪行的严重性以及其他因素进行限定时,他们发现一种显著(尽管很微小)的差异。[24] 黑人法官与这样一种判决相关联:在0—93的严重性量表中,黑人法官的判决要低1.22个单位。当所研究的样本限于黑人被告时,这一影响尤其明显。与白人法官相比,黑人法官对黑人被告的判决,其严重性正好低两个单位——但是,尚有更多可继续讨论的地方。[25]

Probit分析和Logit分析

威尔奇和她的同事还注意到,判决严重性并不是法官量刑裁决的唯一方面,甚至不是最为关键的方面。在对量刑严重性做出判决之前,法官必须决定被告是否将被监禁。一些人受到缓期执行的宽大处理,而且,尽管被判犯有重罪,他们仍未被要求入狱服刑。

当我们以这些方式设想因变量时就无法应用回归分析。相关和回归分析假定,因变量是在区间或顺序层次上得到测量的。然而,判决监禁是这样被编码的:1 = 监禁,0 = 不监禁。倘若情况就是这样,那么,我们就无法讨论自变量一个单位的变化如何引起因变量若干单位数量的变化。

为了处理这些二分的(两类)因变量,另外一种类似于多重回归的分析形

---

[23] *Beta* = −0.01, *b* = −0.48; 不显著。

[24] *Beta* = −0.03, *b* = −1.22。

[25] *Beta* = −0.05, *b* = −2.00。

式已被设计出来。[26] 社会科学研究者通常所处理的因变量就是一些简单的名目分类，比如，“是或否”这样的问题调查；再比如，各种不同的因素如何影响政府决定采纳一项公共政策或加入战争。为了解决这些问题，Probit 分析和 Logit 分析日渐频繁地得到运用。

Probit 并不产生回归系数或贝塔系数，相反，它所产生的系数其自身并不容易得到解释。借助数学公式（简单起见，在此省略），我们可以运用这些系数来评估 X 的变化如何影响 Y 将呈现一个值或其他值的概率。这个统计量有助于回答这个问题，即，法官种族的差异是否影响重刑犯将在监狱服刑的概率。[27]

威尔奇等人利用 Probit 分析来检验自变量和做出监禁判决之间的关系。[28] 他们发现，相比白人法官而言，黑人法官更有可能做出将被告监禁起来的判决。但是，当这项研究的发起者限定其他变量时，这一效应的显著性就会消失。

Probit 分析还表明，曾被限定的因素一旦被引入分析中，那么，在对白人被告做出监禁判决方面就存在显著的种族差异。当白人被告由白人法官判决时，他们更不可能被送进监狱；换言之，当罪行严重性和其他因素被考虑进来时，黑人法官更有可能判处白人被告入狱。

在对所有这些测量进行评估之后，威尔奇和她的同事得出结论：黑人法官的确对刑事司法制度造成影响。以多重回归的结果以及关于这种特殊样本的 Probit 分析为基础，黑人法官比白人法官似乎稍微更有可能：(1)判处白人被告入狱；(2)对黑人被告进行稍微不怎么严重的判罚。但是，“在关于监禁的判决方面，黑人法官似乎是公平的（对黑人被告和白人被告都一样），但是，相比黑人被告而言，白人法官更不可能将白人被告送进监狱”。[29] 请注意，这项研究的发起者对他们的结果进行了限定，并尽力解释多种因素，这些因素可能解释为什么黑人和白人在行为上有所不同。

---

[26] 这种程序指的就是 probit 分析和 logit 分析，或某些情况下所说的“逻辑斯蒂回归”。就因变量的深层分布而言，probit 和 logit 的假定略有不同。参见 J. Aldrich & F. Nelson, *Linear Probability, Logit and Probit Models*, Sage University Paper Series on Quantitative Applications in the Social Science, no. 45 (Beverly Hills, Calif.: Sage Publications, 1984)。

[27] 当相关和回归被用来测定某种关系时，这个统计量可检验一条直线或斜率如何很好地代表数据。当我们运用 probit 分析时，这个统计量检验的是：X 和 Y 之间的关系如何由一条 S 形曲线很好体现出来。寻求判决监禁(Y)和罪行严重性(X)之间线性关联（相关和回归）的统计量，可能会遗漏这种关系，并且导致许多预测误差。

[28] 引自 Welch et al., “Do Black Judges Make a Difference?” 表 1，probit 系数指的是 MLEs（极大似然估计）。Welch 等人认为这个系数是显著的，如果它至少是其标准误差大小的两倍的话（MLE/SE > 2.0）。

[29] Welch et al., p. 134.

表 5.9 变量间关系的测量

相 关	两个区间层次或比例层次的变量之间的关联度或共变度。关系的方向由+或-来表示。
双变量回归	区间层次或比例层次的因变量与单一自变量一个单位的变化相关联的变化量。
多重回归	由若干变量所解释的一个区间层次或比例层次的因变量的变化量。针对每一个自变量的独特影响进行测试。多重回归结合 R^2 一起使用,后者表明由自变量的共同作用所解释的因变量的变化比例。
Probit 分析和 Logit 分析	一种多重回归形式,在其中,因变量是二分的(例如,是/否;赞成/反对)。它考察一个自变量一个单位的变化导致因变量采取一个值或另一个值的概率变化。

我们归纳了到目前为止所提出的有关变量间关系的测量,请参阅表 5.9。

回归的主要问题与下面这一点相关:厘清若干自变量对一个因变量的相互关联的影响或统计上的交叉影响。这个问题通过以下方式得以解决(尽管很少得到圆满解决):精确的操作、关于相似自变量协变异的分析,以及利用像 Probit 分析这样的技术。

在任何一种多重回归模型或 Probit 分析中,大量的技术细节、预防措施以及限制性假定在结果被慎重处理以前必须被考虑进来。但是,我们在此所寻求的就是分析的逻辑。若干变量会影响法院裁决。研究者的逻辑表明,其中一些变量需加以解释或在统计上加以限定,以便他们可以就他们尤其感兴趣的一个变量(种族)提出结论。在这个例子中,情况似乎是这样的:存在一些显著的种族效应,即使我们已经对其他因素(比如,罪行严重性)进行了限定。

尽管统计学使得回归和 Probit 分析似乎技术性很强,但与一般而言的科学一样,它们都始于创造性和想象力。最初一部分的回归分析所涉及的问题,就是弄清楚哪些变量是需要测试的——而这一点又始于对理论的某种意识,以及对所研究的主题的敏锐感受。回归分析的效用就在于,它为我们更为精确地测量假说中的变量间关系指出了可能性。

计算机和统计学

计算机统计软件的发展使研究者得以迅速而又有效地处理数据。一些数学及概念上的背景过去一直是统计学计算的先决条件。现如今,计算机可以进行数学演算。在某些情况下,这也就意味着,我们以与测量层次不相称的方式来操纵数据。

尽管将数学演算留给计算机是可能的,但是,如果不能充分意识到数学处理的概念基础,那么运用统计技术就是很危险的。原因恰恰就在于:虽然一个软件包可以产生出两个变量的相关统计量,但这并不意味着相关所必需的测量标准已经得到满足。我们情不自禁地采用软件所产生的统计量,这些统计量似乎提供了极高的精确性,但是,没有任何东西可替代下面这一点:根据本章(见图5.1)所提出的指导原则,对数据的属性和特征进行慎重的评估。

统计学不创造数据;它们只描述数据。通过指出某一抽象事物的颜色或其他物理属性的方式来描述这一事物,这样做是毫无意义的;同样,声称各种统计关系中没有一种可被计算,也是有误导性的。

我们所讨论的精细改进,其本身只是进一步工作(为了精心制定并且改进研究策略,我们能够做些什么)的开始。我们只是想设法拟定理解和技术的发展方向。通常说来,研究技巧的深入发展,与其说来自贯穿方法论文本的被迫进展,不如说来自一项令人关注的计划所激发的动力。随着这项计划的进展,方法论的重要性就会变得更加显著,并且也更值得学习。

但是,在谋求方法论的理解时,要谨防一种简单的"食谱式"的研究方法。在列举具体的方法之前,一定要弄清楚你想做什么,借助这些具体的方法,你的想法得以实现。这至少是本书的倾向性意见,且也是本书作者的经验。如果能够明了计算背后相对简单的观念,那么,在专门的方法论文献中所发现的众多细节,就会变得更加容易理解。观念提供了技术机制的构架。

引入的概念

测量(measurement)
变量的属性(properties of variables)
名目(nominal)
顺序(ordinal)
区间(interval)
真实零点(true zero)
比例(ratio)
任意零点(arbitrary zero)
概率(probability)
显著性水平(level of significance)
概率样本(probability sample)
分层化(stratification)
抽样偏差(sampling bias)
描述性统计(descriptive statistics)
相关(correlation)
关联方向(direction of association)
相关量表(scale of correction)
列联系数(contingency coefficient)
皮尔逊相关系数(pearson's *r*)
回归分析(regression analysis)
斜率(slope)
多重回归(multiple regression)
标准偏差(standard deviation)
贝塔系数(Beta coefficient)
随机抽样(random sampling)
关联(association)

讨 论

1. 回想下面这一点:样本被用来推测一个更大母体的某些特征。思考这个问题,被用来搜集佩罗数据的抽样技术和方法如何导致结果发生偏差?

2. 仔细查看表5.1(摘自佩罗调查)的问题。在对回答者的态度进行测量时,问题的措辞如何引起误差?

3. 我们能够设想一种用于佩罗调查问题的更为中性的措辞吗?中性措辞怎样并且为什么会造就这样一种测量值,它能更准确地推断出母体的测量值?

4. 一支棒球队的最终分区排名(在分赛区中的位置)是有效阐明球队成绩这个概念的一个区间变量吗?关于球队成绩的更好的区间层次的测量可能是什么?

5. 参阅布罗金顿等人的文章(附录B)中的图1,如果回归系数(或斜率)仅为0.25,那么,代表少数族群人口和少数族群席位份额之间关系的那条线会是什么样子?如果回归系数为1.5,又会是什么样子?较大或较小斜率的实质意义是什么?

第6章 反思：追根溯源

Reflections: Back to the Roots

"与权力结合的知识体现的恰恰就是生命自身的历史……"

——J. P. 狄更斯(John Patrick Diggins)

我们的简要研究始于这样一种求知基础：语言概念从基本的人类经验中的突现。通过对科学操作层面的考察，我们获得一些洞见，借助这些洞见，我们现在可以以一种更为老道的方式来处理建立在这一基础之上的方法结构。是时候该把科学本身置于一种更为宽泛的理解视域之中了。我们有必要对下列问题多一些了解：科学如何与一种更为宽泛的关于知识的观点相契合；科学家如何与科学相关；最后，我们每个人如何能够将科学作为提高我们能力的一种手段来运用，即如何通过科学提高我们对付我们自身环境的能力。

事实性、现实和真实性

科学方法在最初出现时，通常是作为一种狭隘而有限的获得理解的方式。对精确性的严格要求，令人生畏的统计学，经常使得结果难以理解。同时，狂热的科学捍卫者有时很轻率地对科学提出更多要求，其要求已超出科学作为一种知识策略所能承受的范围。他们捏造出一种虚构的客观性，而混淆了如下两个方面：其一就是检验假说的科学程序；其二就是要求个人与专业上能够免除偏差与偏见。

在试图获得关于科学的深刻理解时，我们可以从20世纪一个研究人类状况的伟大理论家身上学到一些东西，这个人就是埃里克·埃里克森(Erik Erikson)。在其作为一个心理分析学家的经历中，埃里克森研究了各种亚文化群，并且广泛地研究了历史上的一些至关重要的人格特征，这使得他最终认为，理解具有多重维度的特征。在我们同我们周围世界的关系中，埃里克森区

分出三种维度:事实性、现实和真实性。[①] 我们将看到,科学同每一个维度都息息相关。

三者当中,事实性最切近流行的科学方法论观点。事实性就是"事实、数据和技术的总体,这些事实、数据和技术可用观察方法和时代的有效技术加以证实"。[②]

我们在此一直在进行的大部分的思索都涉及努力确定那个难以捉摸的研究项——事实。对于事实这个词,我们早先隐约流露出一种个人不满情绪。但是截至目前,我们已经谈论得足够多,目的就是为了澄清这一点:勿将事实与真理混同。某一事实,实际上仅相当于用以确立这一事实的证实方法,也相当于这一事实从中获得意义的参照系。大量的科学都包括,运用方法论的进展来修正、调整甚或证伪那些由较为粗糙或较为僵化的技术业已"证实"的"事实"和理论。我们力图系统地证实观察,通过这种方式,我们强化了我们对世界的感知和外在于我们自身的现象之间的连接。

对深思熟虑的变量说明、精确的测量,以及关于结果的谨慎解释的全部关注,都与发展出那些有资格被称为事实性的数据相关,并且与理解这些数据的局限性相关。尽管事实性世界观似乎很少具有(比如说)理想化观点的魅力和精妙之处,但我们一直试图证明,它有其自身的力量和社会效用。尽管在创造事实性信息上存在的限制,和我们从事该任务卷入的人类特性,都需要一种更为广泛的知识视角,但事实性仍旧是世界观的必要组成部分。

理解的第二个维度或方面是现实,它看似不那么具体,但或许直观上很简单。我们对什么构成现实的感觉,并不仅仅是事实性的叠加。相反,就现实而言,我们所知道的是一种关于事实性的观点,这种事实性由我们藉以理解这些现实事物的感官所整合。只要搜集事实的技术局限性、时间的压力,以及我们理解背景中的无意识因素已固定,那么,我们必须意识到,无论我们多么努力,我们的理解都决不可能是事实性的。而且也无需如此。科学是一门发现并组织证据的学科,这些证据与我们感兴趣的东西相关。然后,我们试图运用那些证据来塑造我们的现实观。因此,我们可以正当地要求,那些致力于科学的人士传达给我们的并不只是"事实",而是他们根据其数据反思得来的关于现实的某种认识。

一门科学要成为社会科学,必须要在科学准则(即,其陈述必须被证实)

---

① 参见 Erik Erikson, *Gandhi's Truth*(New York: Norton, 1969) p. 396; *Dimensions of a New Identity* (New York: Narton, 1974), pp. 33-34; 另见 *Life History and the Historical Moment* (New York: Norton, 1975), pp. 103-104。埃里克森对这些概念的阐释稍微有些差异,我们对其做了修改,以便配合此处说明的需要。埃里克森如何阐发他的分析,关于这一点的一种有趣的说明,可参见 Lewrence Friedman, *Erik Erikson: Identity's Architect* (New York: Scribner's, 1999)。

② Erikson, *Dimensions of a New Identity*, p. 33.

和社会亟需(即,它必须对文明危机有所作为)之间进行一种平衡活动。社会理论的证实通常滞后于社会政策的迫切需要。把被证实的东西与经由一种富有洞见的现实感所获得的思辨结果融合起来,我们就提高了充分理解这个世界的可能性,并且也提高了我们改变世界的可能性。详述这种研究实在的途径绝非易事,而且我们也无法确切说明它如何实现——只能说,个人承诺、经验、将成见搁置起来的意愿,以及良好的科学程序,都会产生某种影响。

然而,事实性(数据和观察的世界)和某种现实感(我们藉以理解证据的观点)尚未构成知识界。埃里克森指出,就我们的目的而言,存在的第三个维度——真实性——意指我们在行动中和通过行动所获得的知识。

科学是反思性研究的生动写照:研究者观察现象以搜集信息,然后归隐在一个安静的地方,汇总、融会并概括所能知晓的东西。然而,这样一种"超然世外"式的理解模式并非我们大多数人的典型模式。人类似乎更趋向行动而不是反思。真实性与我们如何在我们的求知模式以及行为契机影响(或两者之间交互影响)下行动有一定的关联。

通过对他自身作为一个心理分析学家的经验的讨论,埃里克森具体说明了他关于真实性的概念。心理分析主要是一种创造性的行为研究类型。埃里克森认为,治疗从未真正成为这样一个过程:经由这个过程,医生为病人规定某种行动步骤;相反,它是一个相互探索的过程:心理分析学家为探索提供训练和经验,病人则提供他们的个人史、深层情感,以及洞察和行动的能力。心理分析的"交心"治疗与医生和病人之间的潜能相契合。

同样的道理也适用于社会科学研究。我们所研究的行为,并不是被放在载玻片上的,或者只是试管中的泡沫,其形成出于同样的动力法则,这些法则促使研究者作为一个人来进行活动。最好的社会科学家能够献身于他们所研究的行为。他们运用严格的分析,但同时又深入到行动当中,这种行动本身就是一种理解之源。

社会科学家对自身涉入他们所研究的行为这个问题很慎重。显而易见的原因就是,如果不介入由投入所激发的情感,那么有条不紊的思考就相当难以实现。然而,所有社会研究都包括自我与外在于我们自身的某种东西之间的交互作用。作为一种个人立场,"超然世外"与个人涉入一样,有其自身的缺点。无论采用哪种策略,好的研究的确都要求一种相当高的意识水平。科学方法使研究者意识到,他们在处理观察的证实时的(自身产生的)部分交互作用,并且阐明这种交互作用。同样,一个人必须高度意识到,其自身的经验和

个性如何卷入到理解工作当中。[③]

除了迫使我们认清研究中的个人因素之外,埃里克森还指出,实验涉入开启了产生洞见的潜力。行为是反思性的;它通过与环境的交互作用而出现。在某种环境中理解所经验的交互作用需要一种关于何谓人类行为的"感觉"。这种理解需要一种关于事实性的正确评价,以及一种关于现实的观点,而且还需要一种关于行动及行动所能揭示的东西的感觉。

最近,研究者对目前所谓的观察研究的兴趣日渐浓厚。这些研究试图更大限度地掌握尚在研究中的现实。开放式面谈、对环境的引导式描述、个人经验的直接报告,均是这种研究方法的特征。观察研究具有明显的像故事一样的性质,在这种性质背后,通常是一种深思熟虑的努力,以便检验某些理论,并且为假说提供证据。科学与其说是在进行变量说明和关系测量,倒不如说是在搜寻至关重要的观察,并推出它们的理论蕴涵。[④]

每个学生都经历过这种过程:理智地学习某种知识,随后又在经验中重新学习。科学被推崇为一种求知模式,这种模式将最大限度地促进一个人能力的提升,使其能够确定事实、理解事实周围的现实,并且灵活地研究真实性。

科学主要是一种态度和一套普遍方针,而非一种具体策略。有许多可能的研究策略,可在不同层次的事实性、现实和真实性上获得。策略选择是挑战的一部分。[⑤]

女性主义学者已提出新的关于社会科学方法论的观点,这些观点使我们

---

③ 若干社会科学的经典名著的特殊价值应归功于作者个人涉入他们的研究主题。例如 Floyd Hunter, *Community Power Structure* (New York: Doubleday, 1963); Robert Lane, *Political Ideology: Why the American Common Man Believes What He Does* (New York: Free Press, 1962); C. Wright Mills, *The Power Elite* (New York: Oxford University Press, 1959); William Whyte, *Street Corner Society* (Chicago: University of Chicago Press, 1943)。

④ 这些方法的若干资料来源,可参见 Herbert Blumer, ed., "The Methodological Position of Symbolic Interaction," *Symbolic Interactionism: Perspective and Method* (Englewood Cliffs, N. J.: Prentice-Hall, 1969), pp. 1-60; Howard S. Becker, "Problems of Inference and Proof in Participant Observation," *American Sociological Review* 23 (1958); George McCall, "The Problems of Indicators in Participant Observation Research," *Issues in Participant Observation*, eds. George McCall & J. L. Simmons (Reading, Mass.: Addison-Wesley, 1969), pp. 230-239; John Lofland, *Analyzing Social Settings* (Belmont, Calif.: Wadsworth, 1973)。

⑤ 一些资料来源将有助于研究策略间的选择问题: Morris Zelditch, Jr., "Some Methodological Problems in Field Studies," *American Journal of Sociology* 67 (March 1962): 566-576; Donald Warwick, "Survey Research and Participant Observation: A Benefit-Cost Analysis," *Comparative Research Methods*, eds. Donald Warwick and Samuel Asherson (Englewood Cliffs, N. J.: Prentice-Hall, 1973), pp. 189-203; Davydd Greenwood, William Foote Whyte, and Ira Harkevy, "Participatory Action Research in a Process and as a Goal," *Human Relations* 46, no. 2 (1993): 175-192。

在思考证据时有更为广泛的策略选择余地。其研究指出，女人和男人在心理发展中存在显著差异，利用这些研究，这些研究者提出，对人类社会的关系特征的女性主义研究，必须纳入到社会科学的研究机制当中。例如，严格控制数据定义的技术，以及确定问卷回答分类的界限的技术，必须辅以长期的访谈，以及其他形式的定性证据搜集。根据这种观点，激发对人类关系的相互依赖性的研究，无论是通过定性观察还是通过定量观察，都应当成为任何一种社会研究的一个重要组成部分。⑥

道德与科学局限性

先前讨论表明，科学并不回答所有问题，而且科学所确实提供的答案必须被置于其他形式的理解观点之中。换言之，科学有其自身的限度。到了该阐明这些限度的时候了。

对于允许人类以一种文明且和睦的方式共存的道德价值的关注，要求我们接受对下述两个方面的限制：其一就是我们如何进行社会科学研究；其二就是我们可以实现何种结果。对审慎和伦理的关注要求我们，考虑到科学知识所不能证明的东西而对科学知识的主张进行限制，而且要求我们承认，其他知识策略提供更好解答的可能性。以上主题都值得深入研究。

采取一种科学途径来研究人类行为，涉及两类主要的道德问题。在研究计划中对人进行操纵，对于卷入研究的个人而言可能相当危险，而且，科学研究的成果可能被用来剥削人类，而非造福人类。

操纵实验对象所面临的道德困境的一个著名例子，就是研究权威服从的米尔格伦实验。⑦ 这项实验要求自愿者在“科学研究者”的指引下对“学生”实施电击，以便激发他们学会他们正在学习的材料。这些自愿者被告知：这项实验与测试一种方法有关，这一方法可以更有效地教导人们学习某些特定种类的材料。但这项实验的真正要点，其实是测试人们对权威人物（在这个案例中，就是社会科学研究者）的服从。自愿者并不知晓：电击其实是骗人的，而自愿者自身的行为才是这个实验的真正对象。这里包含许多诡计。后来，研

⑥ 参见 Sandra Harding, *Feminism and Methodology* (Bloomington: Indiana University Press, 1987); Carol Gilligan, *In a Different Voice: Psychological Theory and Women's Development* (Cambridge, Mass.: Harvard University Press, 1982)；研究中运用这些观念的一个有趣例子，可参见 Rand Jack and Dana Jack, *Moral Vision and Professional Decisions: The Changing Values of Women and Men Lawyers* (New York: Cambridge University Press, 1989)。

⑦ Stanley Milgram, *The Individual in a Social World* (Reading, Mass.: Addison-Wesley, 1977).

究者向这些自愿者解释这项实验,并为一些自愿者提供进一步的咨询服务。许多自愿者发现他们出于对专家权威的盲目反应,而一直心甘情愿地向人们实施明显危险的电击后都十分沮丧。

若对过度狂热的科学家加以庇护,便有可能出现真正的灾难。一个相当贴切的例子,就是利用未存疑心的人类来实验测定核子辐射和原子弹测试的辐射效应。美国政府目前正努力寻找受害人,并且寻找补救他们的措施。

社会科学的目的应当是改善人类生活质量。这一崇高目的并未证明,运用下列降低人类生活质量的手段就是正当的:无论是通过欺骗使人们做出在正常情况下将感到极其后悔的事情,还是通过揭露他们的内在动机而对带给个人的后果不负责任的方式。社会科学家必须信任实验对象,并且争得他们明确的同意,这是参与的一个前提。

尽管处理实验对象的伦理问题在研究者的控制之下,而且属于研究者的责任,但是,当我们思考社会科学研究的可开发的潜能时,一个更为棘手的道德问题还是会出现。有关社会科学运用的争论因一项涉及那些在20世纪40年代对原子武器进行独创研究的科学家而发生。他们论证说,他们正贯彻科学的路线——科学的运用是其他领域的事情。尽管社会科学家没有原子弹可供炫耀他们的努力,但社会科学已经开始研制的社会控制技术,亦可能被证明具有巨大的影响,这些影响理应受到同样的道德关注。

我们无法在此平息这些道德论争,只是想表明,推动科学进步的原因并不具有特殊的伦理地位。打着科学的幌子做某种事情,不能成为任何人逃避道德考虑的借口,这些道德考虑使人道生活成为可能。

这里所讨论的道德考虑,的确限制着凭借道德良知所进行的研究种类。另外一种限制,并非来自应用于科学的伦理考虑,而是来自科学本身的真正本质。记住,科学开始于(也终结于)不确定性。[8] 科学的任务就是减少不确定性,但最终仍无法消除它。倘若能消除,那科学家便会成为神而不是人。

人们对是否存在神存在分歧,科学无法解决这个问题。科学方法只是一种工具,人类可以运用这种工具,试图减少我们都要遭遇到的不可避免的不确定性。人类本身是能力有限的观察者,而科学所运用的技术和工具亦不完美。这也正是强调外在公开证据和结果可重复性的原因所在。

重点在于,真正的科学家根据现存证据进行概括,但他们不会在证据允许的范围之外做出断言。他们的确不否认,其他类型的知识(比如,信仰、直觉或习俗)也有可能体现智慧,这种智慧超出科学家所理解的证据所及的范围。科学家能够而且应当运用证据技术针对这些其他类型的知识所做的断言进行

---

⑧　这一论点可能扯得太远了。其中一位作者的一位朋友在听到物理学家如何认清宇宙的本质最终是不确定的陈述之后,评论道:“当这些家伙打算炸毁某物时,他们倒似乎非常确切地知道如何做!”

检验。有许多良好的结果可证实这种作为,例如,任何医学史的调查都将证明这一点,然而,科学目前仍无法解释某些医学结果。这些结果可能归因于人类理解范围之外的力量,所有人都可以怀疑这一点,但不能打着科学的幌子否认这一点。以信仰为基础的主张,也不应该被用来为科学必然带来的不宽容进行辩护。

为了使这一论点更接近社会科学,我们来思考心理分析科学的运用和局限。我们知道,成年人的某些伤害行为模式,可追溯至早年遭受的创伤,但这并不意味着,所有行为都发端于孩童时期的经验。显然尚有其他力量在起作用。现代心理分析的创始人,西格蒙德·弗洛伊德(Sigmund Freud)曾注意到,其病人约有三分之一有所好转,三分之一保持原状,还有三分之一病情恶化。并非所有病人的病情都恶化,因为他可能已经提高了三分之一病人好转的可能性,但无论如何这与以完全理解为基础的治疗所能产生的那种结果还是相去甚远。

因此,在解决某些人的问题时,心理分析可能对他们有效,但不能把它作为一个借口,否认其他备选行为解释存在的可能性——至少在证据更为精确之前不能否认。即使到了大部分人类都运用科学的时候,我们身为观察者的局限也不能导致心理分析(或任何其他的科学)做出确定性的主张。更何况,自然科学本来就是在无从测量的无限宇宙中起作用。

此处所提供的信息就在于,对人性价值的道德关注需要对科学的傲慢及对信仰、直觉和习俗的主张进行限制。如果我们打算有效地处理不确定性,那么,对备选理解形式的一定限度的宽容就是必不可少的。若缺少一定限度的宽容,我们便有可能越出我们的人类本性所设定的界限。后果可能相当危险,比如,任何宗教审判、政治大屠杀,以及"科学上正当的"对人类生命的虐待,都可以证明这一点。

科学不是一种道德体系。它是了解生命和宇宙的一种策略,仅此而已。确立信仰、直觉和习俗的界限,超出本书讨论的范围,尽管我们的研究的确建议,所有类型的知识都应当尽力为了文明生活的福祉而和平共存。

我们要了解科学止于何处以及信仰始于何处,除此之外,我们还有必要意识到,除了科学和宗教,还存在其他的研究知识的途径。科学将其自身限定在可观察的事物及我们所称的"现实检验"之中。在这一方面,社会科学家追随自然科学家,后者逐步建立起关于可观察证据的概括。然而,人类行为区别于植物和岩石的地方就在于,它可能由不可观察的力量和预谋所驱使。

举个例子,古希腊哲学家柏拉图认为,我们所看到的"现实"只是感觉印象——感觉印象处在这一过程之中:它靠近或远离隐藏在意见背后的某种完美理念。因此,我们所看到的每张具体的椅子,只不过是椅子理念的不完美实现,椅子的理念只存在于人类想象之中。同样,任何一种现存的政府形式,只

是接近于一种理想化的政府形式,这种理想形式可以从对人类状况的理解中引申出来。这改变了知识的意义。举例而言,根据柏拉图的观点,要"了解"芝加哥政府,就是要观察它合乎一种理想政府类型的地方在哪里。通过与这种理想类型的比较,芝加哥政府的成败得失便可被界定,而且亦可对其未来的表现进行预测。

再举个例子,卡尔·马克思(Karl Marx)表面上为阶级、异化、剥削诸如此类的关键概念提供了太多定义,结果把大多数以科学为导向的读者都搞糊涂了。但是他的所有定义都合乎一种动态模型:人类经由各种形态的历史演变为实现其内在本质而斗争。

人类的内在本质是什么? 马克思通过对人与动物的区分来处理这个问题。我们有能力选择生产什么,而动物则没有这种能力。例如,动物制造蜂房与巢穴,但它们之所以这样做,要么是根据本能,要么是根据偶然的试错法。另一方面,人类可以采集一些树枝来建造居所、小船或一些牙签。马克思相信,当每个人花最少的时间进行必需品的生产,而花最多的时间进行有意识的选择性生产活动时,我们人类将成为真正的人。

这样一来,在马克思看来,剥削可根据不同的方式得以界定,不同阶级的人们在不同的历史时期利用不同的方式,以保护必需品并实现一定程度的独立性。剥削的最终结果是这样一个社会:在其中,所有个体都将分担必需品生产的责任,如此一来,所有人都有可能共享有意识地选择生产活动的自由。它将成为这样一个社会:在其中,没有阶级、没有异化,也没有统治。

柏拉图和马克思所提出的知识形式,在某种意义上超出科学的领域,因为它们都依赖于无法直接观察到的"本质"与"内在关系"。在另外一种意义上,柏拉图体系所容许的预测或者马克思所认同的历史模式在可观察的方面,可以运用科学途径加以检验,这些科学途径可能会阐明它们在解释观察现象时的效用。

我们也有可能进入柏拉图或马克思的方法论世界,并对这些本质和关系的根本假定提出挑战——或者,对马克思关于我们的"类本质"的观点的充分性和完善性提出挑战。⑨ 重点在于,人类有能力超出"是什么"的范围去思考"可能是什么",甚或"应该是什么"。审慎告诉我们:我们需要的是所有知识界中最好的知识,而不仅仅是其中一种完美的知识而已。

---

⑨ 若想深入研究这些观念,可参见 Kenneth Hoover, *Ideology and Political Life*, 3rd ed. (New York: Harcourt, 2001), ch. 6. cf. Bertell Ollman, *Alienation: Marx's Theory of Human Nature*, 2nd ed. (New York: Cambridge University Press, 1976); Paul Thomas, "Marx and Science," *Political Studies* 24, no. 1 (1978): 1-23; Terence Ball and James Farr, eds., *After Marx* (Cambridge, England: Cambridge University Press, 1984), pp. 217-260; and James Farr, "Marx's Laws," *Political Studies* 34, no. 2 (1986): 202-222。

科学家、科学和范式

科学是由人而不是由机器实践的。或者更确切点说，科学是由群体的人实践的。社会科学研究的主要领域都由科学家共同体主导。他们通常定居在大型的研究机构中，并通过期刊、学术会议，以及相互评价和讨论的程序紧密结合在一起。[10] 尽管在这些科学共同体内部，时常存在一些实质性分歧，但对于原则性问题的界限、解决这些问题的标准，以及这些标准的价值（即必须为科学家解决这些问题提供行为规范），科学家通常会达成大致的共识。例如，在美国社会科学共同体中，没有人会撰写独裁政府值得向往的文章。

有一些共同体已涉足社会科学的事业，在评价社会科学研究时，这一事实会带来大量有待考虑的因素。首先，我们很少有人真正愿意独树一帜或与众不同。人们尤其不喜欢必须面对大量的问题，这些问题完全是因他们自己的观点而起。我们的意思是说，在所有人类行为中都存在一种趋于顺从的自然的心理压力，科学研究中同样如此。

若干因素强化了这种趋于顺从的倾向。其中一种因素就是学科的职业建制。尽管多数学生无法察觉，但学术机构中的职位一般都与一种师传相承体制挂钩。师从名师的人，可以谋得最好的职位，也最容易获得一些途径来交流他们的观点。谁能接近老师，以及他或她如何得以将自己的职位发展成为有声望的学术事业，能力当然有很大的关系。但是这套体制的真正影响在于，它极大地强化了固有观点的永恒性，因为徒弟通常会认同师傅的立场。

对于这种顺从的压力，另外一种因素也搀和进来：社会科学研究的政治意义。例如，探查经济系统或社会福利机构中的腐败因素的研究者，不大可能得到他们的目标对象的支持。即便是针对社会经济权力的调查，一旦它成为社区决策的一部分，也会立即引发争论。由于学校和机构通常由代表主导性利益集团的理事会负责运作，所以某些类型的研究计划有可能遭致职业风险。

另外一种影响顺从于较安全的社会解释形式的因素就在于，研究需要经费。优秀的社会科学依赖于调查研究，而这需耗费大量金钱，而且往往需要来自政府机构、企业或基金会的财政支持。能吸引这些资助的专家，不大可能与盛行的社会和政治观念隔阂太远。

基于上述原因，科学研究的特征通常由思想流派或范式体现出来，这些思

---

[10] 关于科学史的一种引人入胜的探讨，详细说明了科学共同体在构造理解时的角色，这一探讨就是 Thomas Kuhn, *The Structure of Scientific Revolutions*, 2nd ed.（Chicago: University of Chicago Press, 1970），此处所讨论的一些主题以这本书为基础。

想流派或范式构成问题界定及解决的方式。然而在面临所有这些压力时,科学方法的最终优点(与其他类型的研究相比)就在于,积累知识的步骤是公开的,且乐意接受质疑与挑战。我们详述这些有损于研究的因素,重点不是要怀疑科学,因为大部分的因素同样会在其他类型的研究中起作用,而是要强调另有原因促使我们对已接受的知识保持批判态度,而对自己的评价标准要保持科学的态度。当我们阅读任何一本书籍、学习任何一门课程,或选择任何一种研究领域时,首先要问的一个问题就是:这种研究类型的背后,主导的范式是什么? 一旦理解了范式,你就有权对证据做出谨慎的评价。

使社会科学满足人类需求

正如我们所看到的,就其作为一种研究工具的全部效用而言,社会科学在其方法论中也带有支配和操纵的可能性。通常说来,研究者运用与人类行为相关的数据,回答的是研究者的问题,而非实验对象本该回答的问题。研究的设计或许将主体转变为被动反应者,其行为以超出主体控制范围的方式得到解释或重新界定。最终,运用研究的目的,可能依赖于有权控制他人的那部分人的优先权,不管该权力是否用于满足人类需求的合法目的。

对于这些问题,以及对于我们在此已描述过的社会科学方法论的若干限制,有一种有趣的研究途径,它已由怀特(William Foote Whyte)、柯林伍德(Davydd Greenwood)、赖兹(Peter Lazes)以及他们在康奈尔大学的同事们提出。[11] 它被命名为参与行为研究,这一技术保留了社会科学研究的精神,同时也以如下方式敞开研究进程:将其效用推广至民众,并提出富有创见的解决问题的途径。以下两个例子可说明如何运用这种技术。

现代复印机的创始者施乐公司,曾因为日本的竞争,一度受到市场占有率急剧下降的威胁。公司的最初反应是考虑将生产作业转移到没有工会的地区,以此降低劳动成本。他们的分析重点是将劳动成本作为关键变量,解决办法很简单:裁员并削减工资。受邀前来协助解决这个问题的一个研究者提出,更主要的问题出在总生产成本上,而不仅仅是占生产成本不及20%的直接劳动成本,而且,对于如何解决这个问题,工人自己或许有一些想法。研究人员、工会会员和管理方共同参与而组成一个"成本研究团队",这使得公司有了广阔的选择余地。其中一个颇富戏剧性的方法是调整间接(非生产)员工相对于直接生产工人的比例,由1979年的2.1∶1改变为1985年的0.4∶1,同时却

[11] 报告见 William Foote Whyte, Davydd J. Greenwood, and Peter Lazes, "Participatory Action Research: Through Practice to Science in Social Research," *American Behavioral Scientists* 32, no. 5 (1989): 513-551。

又使总产量翻倍。这主要是通过如下方式实现的：将监督和控制职能转交给工人本人，同时改变工会章程，以保障不同专业领域中更大程度的就业连续性。⑫

从方法论的观点看，有趣之处就在于：研究"主体"的涉入改变了关键变量的定义，并且改变了所考虑的自变量的范围。成本研究团队并未将注意力仅集中于付给生产工人的工资率，而是着眼于培训、劳动力的连续性，以及间接员工的角色。结果就是：既实现了必要的节约，又恢复了具有竞争力的定价。

第二个例子为讨论增加了新的维度。怀特夫妇对一个超过一百家产业合作社的网络产生了兴趣，它们集中于西班牙北部的蒙德雷格镇。⑬ 这些合作社包括西班牙最大的一家设备生产商，还包括电子元件制造商以及各式各样的在国际上有竞争力的产品。蒙德雷格的制造商合作社网络，已经建立了自己的银行体制、研究机构，以及医疗和教育体系。蒙德雷格的合作社，作为工人所有制，以及在高度精密的生产水平上进行控制的一种模式，已具有国际意义。

合作社面临的一个典型问题就在于，将生产效率与工人/所有者参与的显著性水平协调起来。关键在于如何分析这个问题，以便使蒙德雷格既能够成功，又能允许这种学习得以转移到其他的合作方案中。测量参与的一个常规的社会科学途径，可能涉及对参与意见的调查，另外还牵涉到共同决策的案例分析。以这些观察为基础，就可以发展出与成功的参与形式相关的一般模式。

怀特夫妇及其同事也是运用了这种研究，但他们将之与协商及讨论的过程结合起来，这一过程包括来自不同组织阶层的合作社工人/所有者的圆桌会议。由于圆桌会议不仅要求对参与提出报告，而且还要提出改善参与的建议，所以，先前未料到的参与维度均被揭示出来，并且新的变量可加以概念化和测量。研究的"主体"变成研究设计中的参与者。

上述研究使这些研究者愈加强烈地认识到："……测量受定义左右。糟糕的定义导致误导性的测量，二者结合在一起，又产生误导性的结论"。⑭ 这些研究者通过扩大定义的来源及目的，在工人参与以及集体管理这两个案例中获得了新的洞见。以施乐公司为例，他们最终认识到，有争议的问题远不止工人的生产率；问题在于复杂国际环境下的组织业绩。以就业连续性为导向

---

⑫ Whyte et al., "*Participatory Action Research*," pp. 524-525.

⑬ 参见 *Making Mondragon*: *The Growth and Dynamics of the Worker Cooperative Complex*, 2[nd] *ed*. (Ithaca, N.Y.: Cornell University ILR Press, 1991). 另见 Kenneth Hoover, review of *Making Mondragon in American Political Science Review* 84, no. 1 (1990): 351-352; and "Mondragon's Answers to Utopia's Problems," *Utopian Studies*, 3, no. 2 (1992): 1-20。

⑭ Whyte et al., "Participatory Action Research," p. 548.

开展工作并提高工人的决策参与度,通过这种方式,施乐公司作为一个竞争者的总业绩得到改善。

请注意,重点集中于以行动为导向的分析,而不只是抽象的理解。这种类型的研究通常被称为应用研究,它被认为是纯研究(意指只专注于理智问题的研究)的较为拙劣的同类物。此处的要点在于,在上述案例中应用于行动的研究所产生的概念化和结果,是一个“纯粹”研究者永远无法获得的。这项研究的发起人指出:

> 重新思考过往实践,导致理论的再形成,这相应地又导致实践的改善。重新思考理论和实践的过程,因而同时强化了理论和实践。⑮

参与行为研究具有这样的效应:使本章前面所讨论的三重分析层次(事实性、现实和真实性)全部起作用。上述两个例子的结果就是,这种形式的研究目前已经与一些组织融合,这些组织本身就是适应环境变换的一种方式。

也许过一段时间,你就会受施乐公司之邀来重新组织这个公司,或受邀到西班牙去研究合作社;但是,无论你是在研究学生选举中的参与行为、人们对政治广告的反应,还是在研究人们在工作场所容忍的阶级与地位观念,这种类型的研究都具有应用价值。

科学的激进主义

在讨论完科学机构中的顺从倾向之后,我们最后还要以简单篇幅反向思考,虽然现在便开始讨论科学激进主义似乎有悖常理。确实如此,但我们就这个话题还是有话可说。科学在社会意义上可以是激进的,在个人意义上亦然。

科学研究始于对教条的反抗,这些教条由占统治地位的政治及社会机构所确立和控制。科学史中包含许多关于无畏分析家的重要案例,这些无畏的分析家带着他们的发现走出实验室,其发现威胁到人类研究的各个领域中的主流理解。一些科学家甚至为他们所谓的“异端邪说”付出生命。毕竟,信息控制是政治权力的根基。科学家坚持认为,信息搜集与结论形成程序应当是公开而又可解释的,这样一来,他们便削弱了这样一些人的权力,这些人只会排斥科学研究,而一味地支持“华而不实”的理论,以及自利性的意识形态和学说。

与日常生活更为相关的是一些方法,根据这些方法,科学的思维习惯可以促进你自身能力的提升,使你得以抵制外界条件的作用,并以智慧的方式对付环境。以他人的是非观念为基础、要求我们做这做那的各种理由铺天盖地向我们每个人袭来。多数时候,对大多数人而言,对资源可靠性的评估,足以将

⑮ Whyte et al., “Participatory Action Research,” p. 540.

聪明的建议与无稽之谈区分开。但是,这种区分并不妨碍我们拥有一种独立评价的方式。

长期以来,西方文化一直都将社会问题视为一个关乎人性弱点的问题。这种研究路径,激发了对个人意图、动机和倾向的反思研究。总的来说,社会科学提倡一种不同的途径:事先对你进行全面考察。在决定一个人能否对其行为负全责之前,要考虑环境因素、权力结构、外界的影响、人们在社会环境中所面临的真实的选择维度,以及人们在解决自身问题时实际拥有的物质可行性。与遗传而来的内在倾向,以及有意识与无意识的个体发展史相比,上述那些情况有时更易改变。

科学作为一种方法而进入个人行动之中,这种方法可使我们理解经验的过程有条不紊地进行。科学方法保护措施的存在,主要是为了控制这样一种自然倾向:将我们出于私人目的而希望的世界面貌投射到观察对象之上。尽管科学方法是一种戒律,但在实践中却变成个体解放的一种方法,它使我们从自身观点的狭隘性、自身观察力的局限性、自身偏见的压力中解放出来。在一个亟待改变的世界中,科学(所有人皆可能发展出的一种戒律)可以成为一股激进力量。

科学和政治

改善社会的良苦用心并不缺乏。通常缺少的是一种好的方法。通过国家来强行实现乌托邦,或将其实现留给自私自利的个体的自愿行为,这两种方法都可能带来极大危害。信息充分而又组织得当的理解,可有助于避免两种方法走向极端,并且有助于确立个体间达成共识的基础,而不必使用高压政治或在自私驱使下主张某人的利益高于另一人的利益。

一位敏锐的政治观察家提醒我们:

> 如果政治发展依赖于事实观察,那么虚假的意义就会立即失去它的面纱,而依赖于有效意义的某种共识便会出现,至少在见多识广而又受过教育的观察者当中如此。上述情况并未发生,即使岁月流逝、物换星移。问题、领导者及敌人的一个特征就在于:有关它们意义的争议并未得到平息,正是这一特征,使得它们具有政治性……没有一门政治学涉及这些事务,亦即,呼吁人们就相关事实、它们的意义,以及合理的行动路线达成共识。[16]

---

[16] Murray Edelman, *Constructing the Political Spectacle* (Chicago: University of Chicago Press, 1998), pp. 2-3.

上述这段对政治的批判意味深长,不承认这一点,我们就会曲解这段陈述。只要我们能够以可靠信息为基础解决社会问题,那么,我们就更有希望解决冲突。方法论的训练是这样一种手段:将对信息的歪曲降到最少,同时最大限度地扩大相互理解的机会。

当然,还存在一些本质上具有争议性的问题,这些问题涉及,哪种一致更难达成,如果有可能达成一致的话。[17] 与大多数科学家不同,公民和政治领导人正视行动的必然性。在大多数建设性的政治运用中,人们获得新的洞见并发现一些共同利益,这些共同利益产生出有效的共同体行动模式。本书所描述的在民主背景下实施的方法,可以有助于上述情况的发生,施乐公司和蒙德雷格的例子已表明这一点。政治的领域可以包括,通过社会科学的方法以及其他策略来诚恳地追求真理,尽管某些差异最终须靠政治决策过程加以解决。

政治中还存在另外一种可能。它通过象征性诉求而依赖于意义共鸣,并且依赖于巧妙地运用威胁和再保证,以便调动支持力量,平息反对派。[18] 这种政治的核心是对信息的操纵,并通过信息操纵来操纵民众。另一方面,科学处理信息的方式却可以改善政治。一位德高望重的教师曾发现"科学是一种组织证据的方法——它需要一种社会性的决策过程,这一过程使我们免遭少数人的统治,同时也使我们免遭无知的支配"。[19]

科学和政治都与解决不确定性有关,并且二者都牵涉到真理的证明。毕竟,正是甘地借助一种他所谓的真理力量的政治手段,才打垮了英国在印度的殖民统治。甘地将声讨力量组织起来,这些声讨揭露了英国殖民统治的剥削本质。一旦遭受世界公众舆论的压力,以及被调动起来的民众的新型协同行动,英国便被迫妥协。[20] 美国民权运动的领袖,大都研究甘地的策略,当面临美国法定的种族隔离时,他们也采用了相同的策略。[21] 剥削和统治的真相,一旦通过分析和证明的方式被揭露出来,就被证明要比操纵、欺骗和高压政治本身更具影响。

---

[17] 关于"本质上具有争议性的概念"(essentially contestable concepts)的一种有趣探索,以及它们对政治的意义,可参见 William Connolly, *The Terms of Political Discourse*, 2nd ed. (Princeton, N. J.: Princeton University Press, 1983)。

[18] 相关例子,可参见 Edelman, *Constructing the Political Spectacle*, chs. 3-5。

[19] 这里提及的这位教师就是克劳森(Aage Clausen),在本书写作过程中,他提出很多价值非凡的建议。

[20] 有关真理力量动态学的分析,可参见 Erik Erikson, *Gandhi's Truth* (New York: Norton, 1969)。本章开头关于事实性、现实和真实性的讨论,来自于埃里克森对甘地策略的研究。具体可参阅第396页。另见 Robert Coles, *Erik H. Erikson: The Growth of His Work* (New York: Norton, 1970), pp. 267-399。

[21] Steven Oates 在他的人物传记 *Let the Trumpet Sound: The Life of Martin Luther King, Jr.* (New York: Harper & Row, 1982) 中,回顾了马丁路德·金成为政治领袖的发展历程。

最终，真理力量和社会科学方法论，都依赖于一种对诚实与正直价值的道德承诺。对抗错误和误解的尝试，毫无疑问，必须以追求真理为基础。没有这样一种承诺，政治行动是危险的，科学也将成为一个骗局。

引入的概念

事实性(factuality)
现实(reality)
真实性(actuality)
观察研究(observational studies)
科学共同体(scientific communities)
顺从主义的社会说明(conformist explanntion)
职业建制(career structures)
参与行为研究(participatory action research)
科学激进主义(scientific radicalism)

附录A

调来调去：美国社会资本的奇怪消失[①]

Tuning In, Tuning Out: The Strange Disappearance of Social Capital in America

罗伯特·D. 普特南(Robert D. Putnam)/哈佛大学

……大约过去一年间，我一直在全力对付一个难解的谜。如果我说的没错，这个谜对美国民主的未来而言，具有一定的重要性。它是一个颇费脑力的经典难题，有犯罪事实、布满线索的犯罪现场，以及许多潜在的嫌疑对象。但是，正如在所有好的侦探小说中所描述的那样，一些看似可能的歹徒，结果总有无懈可击的辩解，而一些重要的暗示着预兆性发展的线索，在大幕开启很久以前就已经发生。而且，比如阿加莎·克里斯蒂(Agatha Christie)的小说《东方快车谋杀案》，这一罪行或许本来就有一个以上的罪犯，所以我们必须分清主犯和从犯。最后，我必须首先声明，我也不敢保证我已经解开这个谜。在这种意义上，解谜工作尚在进行。我准备起诉一个首要的嫌疑对象，但证据尚不足以将其定罪，因此我邀请你来帮助厘清线索。

社会资本的理论和测量

几年前，我主持了一项研究，这项研究与意大利地方政府的一个令人费解的主题有关(Putnam,1993)。那项研究的结论是：政府和其他社会机构的业绩，受公民参与社区事务的强力影响，或者说受我所说的社会资本(沿用科尔曼(1990)的说法)的强力影响。我现在正打算应用那套观念和洞见，来分析当代美国公共生活的一些亟待解决的问题。

我所说的社会资本，指的是社会生活的特征——网络、规范和信任，这些特征使参与者得以更有效地联合行动，以谋求共同的目标。当然，他们的共同目标是否值得颂扬，则完全是另外一回事。这些规范、网络和信任将社区的实质性部门连接

---

① 本版本摘自1995 Ithiel de Sola Pool报告，《调来调去：美国社会资本的奇怪消失》，转载自《政治科学和政治学》第664-682页。此处引用已征得剑桥大学出版社的同意。

起来，并跨越潜在的社会裂痕，一旦达到这种程度，即社会资本具有一种“桥梁”作用，那么，增进的合作就有可能满足更广泛的利益，并受到广泛欢迎……

简言之，社会资本指的是社会联系以及随之而来的规范和信任。从这些联系、规范和信任中受益的人——个人、更大的社区或社区内的某一派别——必须在经验上而不是在定义上得到确定②……就当下目的而言，我所关心的社会资本类型，一般说来，是满足公民目标的社会资本。

在这个意义上，社会资本与传统意义上的政治参与紧密相关，但这两个术语并不同义。政治参与指的是我们与政治机构的关系。社会资本指的是我们每个人相互之间的关系。给 PAC 发送支票，是一种政治参与行为，但它并未体现或创造社会资本。在一个团队里打保龄球或者同一个朋友喝咖啡，体现并创造了社会资本，尽管它们不是政治参与行为（群众的政治运动或传统的城市机关，是一种社会资本密集型的政治参与）。我用“公民参与”这个术语，指的是人们与其社区生活之间的联系，而不仅仅指与政治的联系。在一种较狭隘的意义上，公民参与跟政治参与相关，但它们是否保持同步是一个经验问题，而不具有逻辑必然性。例如，某些形式的个体性政治参与（比如开支票捐款）可能处于上升趋势，而与此同时社会联系则在减弱。同样，尽管社会信任（信任他人）与政治信任（信任政治当局）可能在经验上相关，但它们在逻辑上明显不同。我很可能信任我的邻居而不信任市政厅，反之亦然。

社会资本理论假定：通常说来，我们与他人联系越多，就越信任他们，反之亦然。至少就我目前为止所考察的情况而论，这个假定总的来说被证明是正确的：社会信任与公民参与紧密相关。也就是说，无论是否限定教育、年龄、收入、种族、性别等因素，参与（社团或政治）的人总是信任社区或政府的人。③ 而且，这个道理放之四海而皆准，适用于不同国家、美国不同的州、不同的个体、不同种类的群体。④弄清楚因果关系的机制——是参与导致信任，还是信任导致参与——在理论和方法上都很棘手，尽管布莱姆和瑞恩（John Brehm & Wendy Rahn，1995）的报告表明，因果关系主要是参与导致信任。也许是吧，但公民联系和社会信任是协同运行的。它们以何种方式运行？

独舞保龄球：公民参与的趋势

来自大量独立的原始资料的证据强烈地表明，美国社会资本的库存在 25 年多

---

② 在这一方面，我关于社会资本的界定，与科尔曼（James Coleman）关于社会资本的“操作”定义稍微有些不同。参见：Coleman，1990，300-321.

③ 本段以及通篇论文所报告的结果，均引自社会普查，除非另有说明。

④ 从世界价值观调查（World Values Survey）（1990—1991）得来的关于 35 个国家的数据表明，协会会员数的平均数量与支持“多数人可以信任”这一观点的人数之间的相关系数是 0.65。

来一直在缩减。

- 像PTA、慈善互助俱乐部、妇女投票者联盟、红十字会、工会,甚至保龄球联盟这样不同组织的会员纪录表明,许多传统志愿协会的参与,在过去二三十年间大致下降了25%到50%(Putnam 1995, 1996)。
- 在1965,1975,1985年对普通美国人的时间预算的调查(该调查记录了全国范围内抽样的男女在一天中从事的每一项单一活动)显示,自1965年以来,我们花在非正式社交和访问上的时间在减少(大约减少1/4),而我们花在俱乐部和组织上的时间,下降更为明显(大概接近1/2)。
- 尽管在过去三十年间,美国人对政治的兴趣一直稳中有升,而且一些需要动笔的参与形式(比如签署诉状和开支票)大量增加,但大量的集体参与举措却在锐减(Rosenstone & Hansen 1993; Putnam 1996),其中包括,出席集会或演讲(1973—1993年间减少36%),出席与城镇或学校事务相关的会议(减少39%),为政党工作(减少56%)。
- 来自社会普查(GSS)的证据表明,在各个教育层次上以及在男人和女人当中,团体会员数自1974年以来大约下降1/4,社会信任自1972年以来大约下降1/3。⑤ ……会员数的锐减使各种团体感到苦恼,从体育俱乐部、专业协会到文学讨论小组和劳工同盟,无不如此。⑥ ……更有甚者,盖洛普民意测验报告说,做礼拜的人数在20世纪60年代大约下降了15%,而自此以后一直保持在一个较低的水平上,但来自全国舆论研究中心的数据表明,在20世纪七八十年代,下降仍在继续,截至目前总计约30%(Putnam,1996)。

……

为了解释明显的逆向趋势,可能需要对美国社会资本进行一次更为全面的审计。例如,一些观察家认为,互助组和街道巡逻组正在增加,几乎没有人否认,过去几十年间已经见证了利益集团在华盛顿所呈现出的爆炸性增长趋势。……但是,这些并非真正意义上的成员相互会面的协会。它们的会员联系仅限于通用的标志和意识形态,而不是相互之间的联系。……考虑到各种各样的反例,我认为可资利用的有分量的证据证实,当今的美国人和前一代相比,与他们社区的联系显著减少。

当然,大量公民活动依然在我们的社区内随处可见。美国的市民社会并未濒临灭亡。事实上,有证据表明,在社区参与及社会信任的程度上,美国仍高于大多数其他国家(Putnam,1996)。但是,如果我们跟我们自己比,不跟其他国家而跟我

---

⑤ 对政治当局的信任(事实上还包括对多数社会机构的信任)在过去三十年间同样急剧下降,但这种下降在概念上是一种截然不同的趋势。我们从下文将看到,社会信任下降的原因和政治信任下降的原因完全不同。

⑥ 出于下文所解释的原因,图1报告了各种类型团体中的会员数的趋势,条件是限定应答者的教育水平。

们的父辈比，最有力的证据都表明，我们彼此间的联系在减少。

本序言提出大量值得深入探讨的重要问题：

- 美国社会资本的库存真的减少了吗？
- 它要紧不？
- 我们能够对它做些什么？

在我看来，对前两个问题的回答是肯定的，但我无法在此深入处理它们。最终使我最为关心的是对第三个问题的回答，这一答案至少在某种程度上首先取决于，我们对扰乱美国公民生活的奇怪弊病的原因的理解。这正是我打算在此解开的那个谜：为什么美国社区生活的组织结构，在20世纪60年代开始削减，并在七八十年代进一步加剧？为什么越来越多的美国人独自打保龄球？

对社会资本减少的解释

这个谜题有多种可能解答：

- 繁忙和时间压力
- 经济上的难关（根据一些备选理论的观点，或者是物质富裕）
- 住所机动性
- 郊区化
- 妇女成为有偿劳动力以及双职工家庭的重担
- 婚姻和家庭关系的破裂
- 美国经济结构的变化，比如连锁店、分公司和服务业的增加
- 20世纪60年代（大多数实际上发生在20世纪70年代），包括

——越南战争、水门事件，以及对公共生活的理想的幻灭

——反抗权威的文化叛逆（性、毒品等）

- 福利国家的增多
- 民权革命
- 电视、电子革命及其他技术变革

大多数有名望的侦探作家，都不愿意清点这么多看似可能的嫌疑对象，无论虚构的侦探多么能干。说实话，我无法处理上述所有理论——当然是指不能以权威的方式处理，但我们必须开始筛选这个清单。可以肯定，与我们正在研究的那种趋势一样普遍的某种社会趋势，可能有多重原因，因此我们的任务是评估这些因素的相对重要性。

解开我们谜题的某种途径，即使不全面，也须通过若干测试。

所提出的解释因素是否与信任及公民参与相关？如果不相关，就很难明白，为什么这个因素还被列入这个清单中。例如，许多妇女在我们尚在讨论的那一时期，

已经成为有偿劳动力,但是,如果职业女性被证明比家庭主妇更投入社区生活,那么,就很难将社区组织中的下降趋势归因为双职工家庭的增多。

这种相关是虚假的吗? 例如,如果有子女的家长比无子女的人更有可能成为参与者,那么这将成为一个重要的线索。但是,如果家长身份与公民参与之间的相关被证明是完全虚假的,例如,由于年龄的影响,我们将不得不把下降的出生率从我们的怀疑清单中剔除。

所提出的解释因素是否以相应的方式发生改变? 例如,假设经常搬迁的人具有较浅的社区根基。那么,只有当住所机动性本身在这一时期上升时,上述假设才能成为解答我们的谜的答案的一个重要部分。

所提出的解释因素可能会遭受如此诘难吗?即它可能是公民不参与的结果,而非原因。 例如,即便报纸读者群跨越个体和时代限制,与公民参与紧密相关,我们也有必要权衡下列可能性:报纸发行量的削减是不参与的结果(而非原因)。

依据上述那套基准,我们来考虑影响社会资本形成的各种潜在因素。

教　育

人力资本和社会资本紧密相关,因为教育对信任、协会会员,以及许多其他形式的社会和政治参与,均具有强烈影响。时至今日,教育是我所发现的各种形式的公民参与(包括社会信任以及许多不同类型的团体中的会员数)的最牢固的相关项。[⑦] ……简言之,教育是预测公民参与的一个有力指标。……

在某种程度上,教育代表了社会阶层和经济的差异,但是,当收入、社会地位和教育被同时用来预测信任和团体会员数时,教育仍是首要的影响因素。……简言之,受教育程度高的人,更有可能成为参与者与信任者,部分原因在于他们经济上更富裕,但主要还是因为他们在家及在学校所接受的技能、才略与爱好。

人们普遍认为,当今的美国人要比我们的父辈和祖辈受到更好的教育。然而却很少意识到,这种趋势在刚刚过去的20年间如何对成年人的教育构成带来巨大而又迅速的改变。自1972年以来,所有成年人中受过12年以上教育的人所占的比例几乎翻倍,从28%上涨到50%。……

这么看来,教育急剧地推进了公民参与,教育水平也大幅度提升。不幸的是,这两个无可否认的事实,只会加深我们的主要疑惑。照理说,在过去20年间,教育水平的提升本应该增加社会资本……但与之相反,自20世纪70年代初以来,实际的GSS数据表明,社会资本呈净下降趋势。……

因此,上述第一种研究尝试加深了我们的困惑。尽管如此,我们还是可以从上

⑦ 唯一的例外是庄园集团、劳工同盟和退伍军人组织,它们的成员所接受的正规教育比普通美国人要稍微少一些。

述发现中得出两个有用的结论，一个是方法论上的，一个是实质性的：

1. 既然教育对公民参与和社会信任有如此强烈的影响，那么，我们在考察其他的可能因素时，就有必要说明教育的差异，目的是要确保，我们不会把教育的结果与其他变量的可能影响混淆起来。
2. 无论公民参与和社会信任骤降的背后存在何种力量，这些力量已经影响到美国社会的各个层面。过去25年间，令人费解的不参与似乎已经使我们社会的所有阶层都感到苦恼……

时间和金钱的压力

当今的美国人的确觉得比上一代人更忙碌：20世纪60年代中期至90年代中期，我们这代人中报告说感觉“总是很匆忙”的人数比例大幅度上涨（Robinson & Godbey，1995）。我们倾向于逃离社区事务，或许，隐藏在这种倾向背后的、最明显的嫌疑对象，就是到处充斥的繁忙。而且，潜藏在幽暗处附近的，正是那些我们今天已大量讨论过的我们社会所特有的经济压力——工作不安全感和实际工资的下降，尤其是在收入处于低2/3部分的人当中。

然而，尽管繁忙和经济上的不安全感乍一看似乎应受到谴责，但我们很难找到将其定罪的证据。事实上，证据的天平有力地显示：对我们设法要解开的那个谜题而言，时间和金钱压力显然不是重要的促成因素。

首先，时间预算研究并未证实这一论点：即平均说来美国人比上一代人工作时间更长。相反，罗宾逊和戈德贝（Robinson and Godbey，1995）报告说，1965—1985年间，普通美国人平均每周多赢得5小时空闲时间，部分是因为花在家务上的时间减少，部分因为较早退休。他们宣称，当今美国人比几十年前拥有更多空闲时间，毫无疑问，这一论断会遭到其他观察家的质疑。例如，绍尔（Schor，1991）的报告证实，我们的工作时间在加长，尤其是对妇女而言。然而，不管上述争论的解决途径是什么，仅仅凭借工作时间与社会信任及团体会员之间的相关性考察，就将公民不参与归咎于工作日的延长，这个论点更不合理。

可资利用的证据强烈显示：事实上，长时间的工作与公民生活参与的减少或社会信任的减少并没有关联。完全相反：来自社会普查的结果表明，相比那些不在有偿劳动力之列的人而言，就业一族加入的团体还稍微要多些。尤为令人震惊的是

这一事实:较长的工作时间与更多的(而非更少的)公民参与相联系。[8] 这一惊人发现,与来自时间预算研究的证据完全符合。罗宾逊(Robinson,1990a)报告说:不足为奇,花更多时间投入工作的人的确感觉更匆忙,这些饱受折磨的人,在吃饭、睡觉、读书、从事业余爱好,甚至无所事事方面,的确只花更少的时间。与其他人相比,他们花在看电视上的时间少得更多——几乎少了30%。然而,他们花在组织活动上的时间并没有减少……

……而且,全国范围内的参与和信任的下降,在全职工作者、兼职工作者,以及在那些不在有偿劳动力之列的工作者当中,均有相当体现。因此,即使人们正摆脱社区生活,延长的(工作)时间似乎也不能成为理由。

如果时间压力不是我们要寻求的根源,经济压力又会如何?……相比穷人和中层收入的工薪族而言,在更富裕的一部分美国公民中,参与和信任方面的下降实际上还略微严重一些。……总之,无论是客观的还是主观的经济安乐,均不能使美国人预防公民不参与的病毒;甚至正相反,富裕在一定程度上加剧了这个问题的恶化……

流动性和郊区化

大量研究已经发现,住所稳定性以及相关现象(比如,私房业主)与更多的公民参与相关。在调查研究的早期阶段(Putnam 1995, 30),我发现"与植物的频繁移植一样,流动性往往会破坏根部系统,而一个背井离乡的个体要种下新的根基需要花费时间"。但我现在必须报告说,深入的调查完全可以证明,住所流动性不应当对公民参与的下降负任何责任。来自美国人口普查局1995年(和早些年)的数据显示,住所流动率在过去半个世纪一直相当平稳。事实上,非要说存在任何变化的话,也只能说长短距离的流动在过去50年间均呈下降趋势。在20世纪50年代,每年有20%的美国人变换住所,6.9%的美国人跨州迁移;在90年代,同比数字分别为17%和6.6%。总之,当今的美国人相比前一代人而言,住所更稳定。如果有关经济萧条解释的定论还得更精确的话,那么,关于流动性的定论是一清二楚的。这个理论就是错的。

但是,如果迁移本身并未侵蚀我们的社会资本,那么下列可能性又如何:我们

---

[8] 无论是否限定教育和调查年份,这一事实都是对的。在男性和女性当中,关于这个分值的模型是不同的,因为兼职妇女比全职或足不出户的妇女,在公民参与和社会信任方面,稍微更积极一些。不管我们怎样对待这种让人感兴趣的反常现象,它显然并未出现在时间预算数据中(Robinson & Godbey,1995),而且在男性当中,情况亦然,这种反常无法解释我们基本的谜题,因为兼职女性在美国总人口中所占的比例相对较少,而这一比例在增长,而不是在下降。GSS数据显示,20世纪70年代前半期至90年代前半期,兼职女性在整个成年人口中所占的比重,由近8%上升至近10%。

一直往那些更不适宜社会联系的地方(尤其是郊区)迁移?为了检验这一理论,我们必须首先研究居住地和社会资本之间的相关性。事实上,社会联系的确因社区类型而异,但这些差异被证明是适度的,而且与上述理论背道而驰……

妇女角色的转换

大多数(我这一代人的)母亲都是家庭主妇,而且她们大都在社会资本形成上有很重的投资。这里所说的投资是一种专门用语,指的是在教会晚餐、PTA 会议、邻里咖啡聚会,以及拜访亲朋好友等方面所耗费的不计其数的、不计报酬的时间。妇女离家并成为有偿劳动力,这一潮流也许是过去半个世纪最不寻常的社会变化。无论女权革命多么受欢迎或多么过时,我们都很难相信它未曾对社会联系产生任何影响。难道这是过去 25 ~ 30 年间社会资本下降的首要原因?

可资利用的调查证据中,某些模型似乎支持这一主张。考虑到各种因素,女人从属的自愿协会比男人稍微少一些(Edwards, Edwards & Watts,1984)。另一方面,时间耗用研究表明,相比男人而言,女人在那些团体以及非正式的社会联系上要花去更多时间(Robinson & Godbey,1995)。尽管参与和信任方面的绝对下滑在男人和女人当中几乎相等,但相对下滑在女人当中稍微严重一些。限定教育因素,那么,男人当中的会员数每十年以 10% ~ 15% 的比例下降,而女人当中的同比下降比例为 20% ~ 25% ……

但是,诚如我们先前所见,工作状况本身对团体会员数或信任所产生的净影响似乎很小。相对于职业女性而言,家庭主妇从属于各种不同类型的团体(例如,更多的 PTA,以及较少的专业协会),但在总数上,职业女性实际上自愿加入的协会要稍微多一些。[9] 更为重要的是,公民参与的总体下滑在家庭主妇中要比在职业女性中更严重。1965 和 1985 年的时间耗用数据的对比(Robinson & Godbey,1995)似乎表明,职业女性作为一个群体,在组织上所花的时间较之以前实际上更多,而非职业女性花的时间则更少。同样的研究表明,自 1965 年以来,非正式社交中的主要下滑也集中在无业女性当中。当然,主要事实还在于:对各类女性而言,总趋势是下降的(对男人而言同样如此,甚至还包括单身汉),但数据显示,相比非全职工作的妇女而言,全职工作的妇女对这种下滑趋势实际上更具免疫力。

如此看来,妇女似乎应当对公民参与在过去 20 年间的下降承担一定的责任,尽管如此,我们还是很难找到微观层次上的数据,将上述事实与妇女加入劳动力大

---

⑨ 但是,罗宾逊和戈德贝(Robinson and Godbey,1995)报告说,相对于职业女性而言,非职业女性仍然会花更多时间来参与自愿性协会的活动。

军联系起来。当然,我们很难控制这些数据中的选择偏差,因为那些已经决定加入劳动力大军的妇女,毫无疑问,在许多方面不同于那些已经决定居家的妇女。社区参与之所以在职业女性当中似乎在增加而在家庭主妇当中似乎又在下降,其中一个原因或许就在于:正是那些在早些年代最频繁参与社区事务的妇女,最可能加入劳动力大军,其结果同时也就降低了公民参与在剩余那部分家庭主妇中的平均水平,而同时又提高了职业女性公民参与的平均水平。很明显,一直以来我们并未进行一项重大的全国性的控制实验,以研究工作对女性公民参与的影响,而且不管怎么说,数据模型尚不完全清楚。但是,与我本人的早期推测相反,我几乎找不到任何证据支持这个假说:在过去的一代中,妇女加入劳动力大军的潮流,已经对社会联系和公民参与产生了重要影响。另一方面,我也找不到任何明晰的替代性说明以解释这一事实:公民参与的相对下滑,在女人当中要比在男人当中更严重。既然这种证据至多是间接性的,那么,我们在此所能作出的最好的暂行判断,也许就是著名的苏格兰式裁决:证据不足,暂不定罪。

婚姻和家庭

另外一种广受讨论的、与公民参与的下降多少有点关联的社会趋势,就是传统家庭单位(妈妈、爸爸和孩子)的瓦解。根据某些解释,既然家庭本身是社会资本的一种关键形式,那么,也许它的消逝是参与和信任在较大社区内下降的部分解释。证据表明什么?

首先,家庭纽带松弛的证据是显而易见的。长达一个世纪的离婚率的上升(20世纪60年代中期至70年代中期加速增长,随后趋于平稳),单亲家庭近来也不断增多,除此之外,单人家庭的发生率自1950年以来已经翻了一倍多,部分原因在于单独生活的寡妇数量的上升(Caplow, Bahr, Modell, and Chadwick 1991, 47, 106, 113)。正如社会普查所反映的那样,所有这些变化的净效应就在于:所有美国成年人中,目前尚未结婚的人数比例由1974年的28%,攀升至1994年的48%。

其次,已婚男女在我们关于社会资本的两种测量中,的确处在稍高的等级上。也就是说,限定教育、年龄、种族等因素,单身族(离婚的、分居的,以及从未结婚的男女)比已婚族明显更不信任他人,也更不热衷公民参与。[10] 大致说来,与单身男女相比,已婚男女的对外信任度大约高1/3,而他们所从属的团体大约多15%~25%。(在这项比较中,寡妇和鳏夫更像已婚族而非单身族)

⑩　多重变量分析有迹象表明,离婚使社会联系下降的一个主要原因就在于,它降低了家庭收入,这反过来又降低了公民参与。

总而言之，成功的婚姻（尤其是当这个家庭单位还包括孩子时），在统计上与更多的社会信任和公民参与相关。因此，信任和会员数的部分下降，与婚姻率的下降有关……

我本人的断定是（一定程度上以下文引入的额外证据为基础），婚姻的破裂或许是本文开头提到的那宗罪行的从犯，而非主犯。

福利国家的兴起

在一些评论家看来（例如，Fukuyama，1995，313-314），间接证据，尤其是关于社会联系下降的定时统计，已经表明，公民不参与的一个重要原因（甚或正是这个原因）就是大政府和福利国家的壮大。据称，通过对个人主动性的“驱除”，国家干预已颠覆了市民社会。这个题目太大，我无法在此详述，只是提出一两点适当的看法。

一方面，几乎可以肯定，某些政府政策已造成社会资本的破坏。例如，20 世纪 50 年代和 60 年代的所谓的“贫民窟清除”政策，切断业已存在社区纽带。还可以设想，一些社会支出和税收政策，可能已经对热心公益的慈善事业造成了障碍。另一方面，我们又很难看出，哪些政府政策应该对保龄球联盟和文学俱乐部的衰落负责。

解决上述问题的一种经验途径，就是考察公民参与的差异，以及跨不同政治辖区的公共政策，以便弄清楚臃肿的政府是否会导致社会资本的缩减。然而，在美国各州，社会资本的差异与各种关于福利支出或政府规模的测量，表面看来并没有本质的关联。⑪ ……

种族和民权革命

种族是美国社会史的一个绝对根本的特征，这个社会的几乎所有其他的特征都以某种方式同它相关联。因此，我们似乎可以在直觉上合理地认为，种族对持续一代之久的社会资本的削减，可能在一定程度上产生了影响。事实上，一些观察家（包括黑人和白人）已经注意到，恰恰就是在 20 世纪 60 年代的民权革命取得最辉煌的成功之后，社会联系和社会信任开始下滑。在某些观察家看来，上述巧合已经

⑪ 我会在以后的著作中更详尽地处理这个问题。但对以下纪录，我要着重指出：（1）社会信任和团体会员数的国家层面上的差异，是实质性的、紧密相关的，并且是适度平稳的，至少 20 世纪 70—90 年代这一时期是如此；（2）根据 Elazer（1966）所发明的关于“国家政治文化”的测量，上述那些差异惊人地紧密相关（$R^2=0.52$），Sharkansky（1969）基于对整个 50 年代的国家政治的描述性说明，对 Elazer 的测量进行了改进，而这种改进反过来又可追溯至 19 世纪及其以前的移民模型。

间接表明了一种社会学意义上的"白人逃亡"的可能性,因为公民生活中法定废除种族隔离,致使白人从社区协会中退出。

与关于福利国家的理论一样,上述对社会资本破坏的种族性解释相当有争议,仅靠上述简短评论,几乎无法解决这一争议。尽管如此,基本的事实还是这些。

首先,协会会员间的种族差异并不大。……而另一方面,种族差异在社会信任方面的确相当大,即使考虑到教育、收入等的差异,情况亦然。平均说来,在1972—1994年间,限定教育差异,大约有17%的黑人支持"多数人是可信的"这个观点,相比而言,大约45%的白人赞同这一观点,大约27%的其他种族的人认同这一观点。[12] 当然,社会信任方面的这些种族差异,反映的并不是集体偏执,而是贯穿多代人的真实经历。

其次,社会资本的削减已影响到所有种族。……尤为重要的是,不参与在白人中的发展速度,与种族互斥或对种族隔离的支持,一直毫不相关。相比更宽容的白人而言,公然的种族主义者或主张隔离主义的白人,在这一时期并没有更快地脱离社区组织。……这个证据还表明,废除过去30年的民权进展,还是无法挽回社会资本的损失。

代际效应

我们一直在竭力找出公民不参与的根源,可我们的努力目前为止并未取得突出进展。下降趋势一如既往地贯穿美国社会的主要领域。……然而,一个明显的例外涉及年龄。在我们所有的统计分析中,作为所有形式的公民参与和信任的一个预测因素,年龄仅次于教育。老人比年轻人从属于更多的组织,并且他们很少会愤世嫉俗。而且,年长的美国人更经常地参与投票,更频繁地读报,这是与参与和信任紧密相关的另外两种形式的公民参与。……大多数评论家都将这种模式解释为一种生命周期现象,我最初也这么认为。

来自社会普查(GSS)的证据使我们能够随着个体的老化来了解他们的年龄组。……然而,令人吃惊的是,这种分析(不断地被用于研究在各个连续的年龄段出生的个体)并未提供任何有效证据证明公民参与中的生命周期变化。事实上,更常见的情况是:当几代人都经历了1972—1994这一时期时,他们的信任度和会员数不升反降,这种情况反映出,公民参与在年轻人以及在老人当中存在一定程度的同时性下滑,80年代后半期尤为明显。但是,上述下降趋势明显不能解释,为什么经历过那一时期的年长美国人,信任度和参与度往往会更高。事实上,根据这些数据,我们唯一可发现的可靠的生命周期效应就是:当我们走过我们自己的80年代时,我们会在垂暮之年退出公民参与。

---

⑫ 与本文别处的用法一样,这里所说的"限定教育差异"是指,对少于12年、正好12年,以及多于12年学龄的应答者的平均分值,分别进行平均化。

这些模型所暴露的一个主要悖论是这样的：老年人比年轻人一向都更积极地投入公民参与，而且也更信任他人，然而，随着我们逐渐老化，我们并未变得更积极，也更不信任他人。到底出了什么问题？

时间和年龄对社会行为的影响相当模糊。社会科学家已经认识到，要区分三种截然不同的现象：

1. 生命周期效应所体现的差异，可归为不同的生命阶段。在这种情况下，个体随着年岁增长而发生改变，但是，出生和死亡的"人口统计上的新陈代谢"，使老化效应总体上趋于匀称的平衡，因此，生命周期效应并未发生总体改变。随着年岁增长，我们每个人的近距视力都会下降，但对放大镜的总体需求几乎没有变化。
2. 时期效应对经历过某一特定时代的所有人均产生影响，不管他们的年龄如何。[13] 时期效应既能引起个体改变，又能引起总体改变，这种影响通常迅速而又持久，并且无视与年龄相关的差异。例如，1965—1975 年间，人们对政府的信任急剧下滑，这种趋势基本上属于这种时期效应，因为各个年龄段的美国人，都改变了他们对其领导人可信度的看法。同样，正如我们刚刚所注意到的那样，社会资本在 20 世纪 80 年代所出现的适度下滑，似乎也是一种时期效应。
3. 代际效应，正如曼海姆（Karl Mannheim）在其经典论文《代的问题》（*The Problem of Generations*）中所描述的那样，代表着这一事实——"属于同一代人的个体，只要出生年份相同，那么，他们在社会进程的历史维度中，就会被赋予共同的定位"（Mannheim，1952，290）。与生命周期效应一样（不同于典型的时期效应），代际效应表现为年龄群组在某一单一时间点上的差异，但与时期效应一样（不同于生命周期效应），代际效应引起真实的社会改变，因为，被深深打上不同观点"烙印"的一代又一代的人，总是你方唱罢我登场，接连不断地登上历史舞台。在纯粹的代际效应中，个体从未改变，但社会却发生了改变。

……回到我们的难题，倘若年长的人想当初并未随着年岁增长而变得更积极地投身于公民参与并且更信任他人，那么，他们今天又如何会变得更积极而又更信任他人呢？正如巴特勒和斯特克斯（David Butler & Donald Stokes，1974）在另外一种语境下所注意到的那样，解决这个悖论的关键，不是追问人们现在多大年龄，而是追问他们何时年轻。图 1 再现了这个问题，它表明了根据被调查者的出生年份所进行的各种关于公民参与的测量。[14] ……

---

[13] 只对特定年龄段的人产生影响的时期影响会逐渐变成代际效应，这正是康弗斯（Converse）在概括与年龄相关的效应时，仅提到"两个半"类型而非传统的三种类型的原因所在。

[14] 为了排除生命晚期中的生命周期效应，图 1 不包括 80 岁以上的调查对象。为了避免在对青年人进行可靠抽样时所存在的众所周知的问题，康弗斯（Converse，1976）曾就这些问题做过讨论，图 1 同样不包括 25 岁以下的调查对象。为了弥补相对较小的逐年样本的不足，并且为了限定教育差异，图 1 绘制出五年移动平均线，它们横切本文所运用的三种教育范畴。

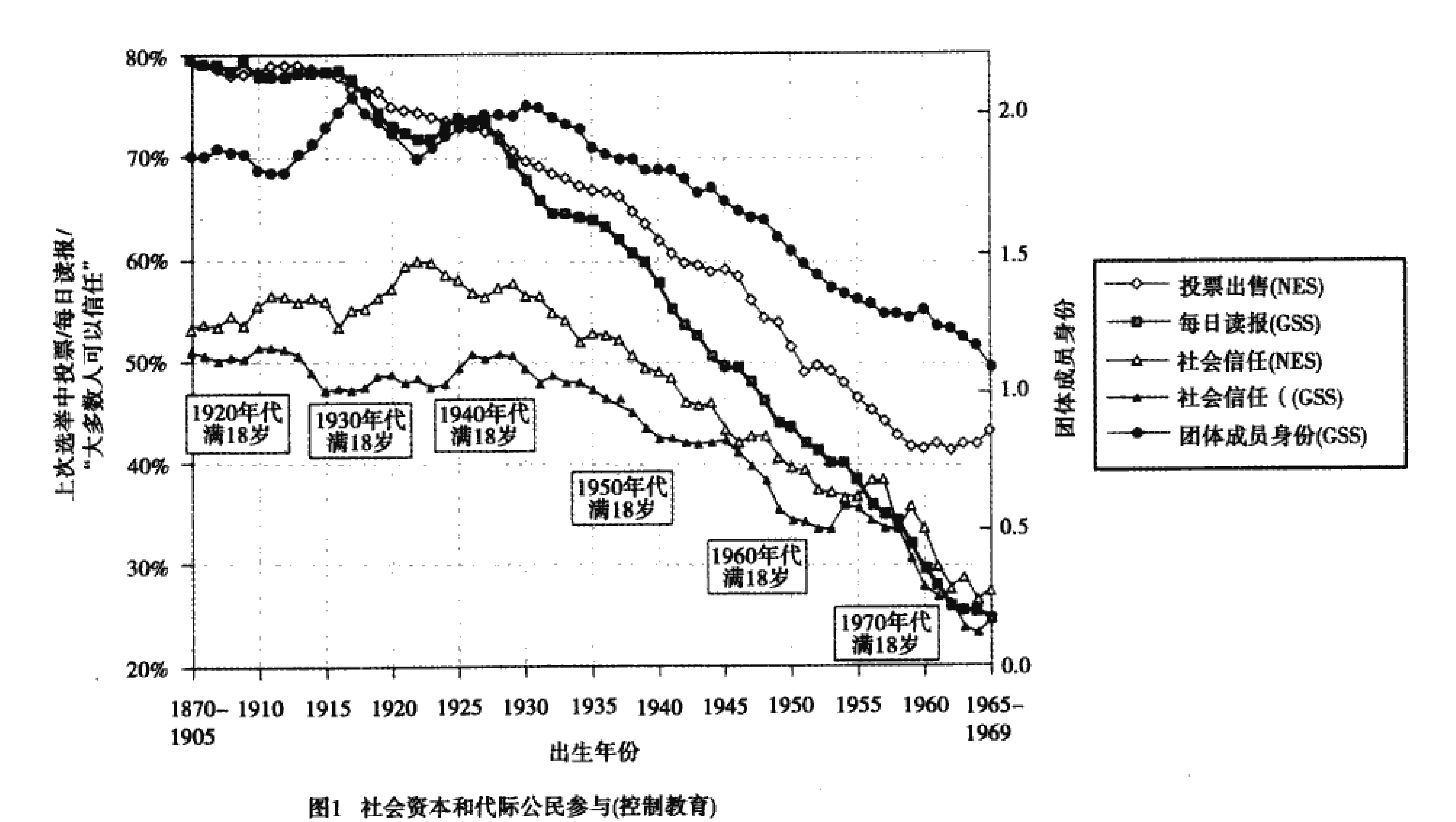


图1　社会资本和代际公民参与(控制教育)

来源：General Social Survey (GSS), 1972–1994 and National Election Studies (NES). 1952–1992

回答者年龄为25~80岁，平均5年一段。

三类教育范畴平等加权。

漫长的公民代

实际上，图 1 从左至右按照出生日期将美国人排成一线：起点是那些出生在 19 世纪后 30 年的人，并且一直延续至他们的曾孙一代，也就是 20 世纪最后 30 年出生的一代人。当我们沿着这条横轴从左至右（从那些在世纪之交长大的人，到在兴旺的 20 年代成长的人，以此类推）移动时，我们发现，公民参与和社会信任保持一种相对较高且又不平稳的上升趋势。然而，我们接下来就看到社会参与相当突然地减少的迹象，始于出生于 20 世纪 30 年代早期的男女。值得注意的是，参与、信任、投票及读报中的这种下降趋势，几乎不间断地持续了近 40 年。公民参与的各种不同指标的轨迹惊人地平行：对于出生且成长于 20 世纪前 30 年的人而言，每一轨迹都表明一种高的、有时是上升的稳定态势；20 世纪 30 年代左右出生的年龄群组中，每一轨迹都呈现出一个转折点；而且，每一轨迹随后都表现出程度不一的持续下降趋势，一直延续至 60 年代出生的年龄群组。⑮

不管根据哪种标准，这些代际差异都是异乎寻常的。举个例子，我们来比较出生于 20 世纪 20 年代早期的一代人和出生于 20 世纪 60 年代的他们的孙子辈的一代人。限定教育差异，那么，出生于 20 世纪 20 年代的一代人，加入公民协会的会员人数几乎是出生于 20 世纪 60 年代的那些人的两倍（前者人均会员数约为 1.9，后者约为 1.1）。祖父辈更有可能信任他人，其中信任他人的人数比例是他们孙子辈的两倍多（50% ~60% 对 25%）。他们参与投票的人数比例几乎是最近年龄群组的两倍（约 75% 对 40% ~45%），而且，他们读报的人数几乎是通常读报人数的 3 倍（每天都读报的比例是 70% ~80% 对 25% ~30%）。记住：我们尚未找到任何证据，证明最年轻的一代人随着年岁增长，最终将与他们祖父辈较高程度的公民参与相匹配。

因此，对我们的基本谜题而言，我们的数据中与年龄相关的模型提出了一种完全不同的解释：不能将这些模型理解为生命周期效应，而应理解为代际效应。用这种方法解释，图 1 描述了一个漫长的“公民”代，这代人出生于 1910—1940 年间，这是一个广泛的群体，比起年轻一代，他们更多地参与社区事务，而且从根本上也更

⑮ 在社会普查的数据中发现公民参与的代际差异之后，我们才获悉了米勒/夏克斯（Miller/Shanks）的论点，但他们的发现与我的发现惊人地一致。

信任他人。[16] 这一公民代的顶峰是出生于1925—1930年间的年龄群组,他们在美国经济大萧条期间上小学,在高中(或在战场上)经历了第二次世界大战,在1948年或1952年首次投票,在20世纪50年代成家立业,并且在二三十岁以后才首次看电视。由于全国调查已开始,这代人一直格外具有公民精神:更多地投票、更多地参与、更多地读报、更多地信任他人。正如著名的社会学家蒂利(生于1928年),在评论本文的早期版本时所说的那样,"我们是最后的乳儿"。

为了有助于解释代代相连的美国人在其中成长起来的历史背景,图1还指出了每一年龄群组在其中得到法定年龄的十进制期限。因此,我们可以看出,20世纪40年代后成年的每一代人,均比其相邻的前一代人,更少地参与公民事务。

对这种代际解释的进一步证实来自两条平行线的比较,这两条平行线在图中均表示对同一个社会信任问题的回应,最初是在全国选举研究(NES)(主要在1964—1976年间)中被提出,随后又在1972—1994年间的社会普查(GSS)中被提出。[17] 如果生于20世纪初的美国人所表现出的更高程度的信任代表一种生命周期效应的话,那么来自GSS调查(当时被调查对象的平均年龄为10岁)的曲线,本应该在NES线条之上间隔一定距离。实际情况却是,GSS线条大约比NES线条低5%~10%。几乎可以肯定,这种向下的转变代表一种时期效应,它导致社会信任在20世纪80年代的各个年龄组的人群中都呈下降趋势。[18] 但是,这种下降的期间效应,比起我们已注意到的巨大的代际差异而言,实质上要适度得多。

简言之,对公民参与中与年龄相关的差异所做的最简洁的解释就是:这些差异代表了公民参与在二战后的几十年间成年的美国人当中的一种急剧的下降,而且还代表着一定程度的不参与,这些不参与影响到20世纪80年代的各个年龄段的人。这些模型提示我们,在二战后成长与在水门事件之前成长,是一种完全不同的经历。战后的各代人似乎被暴露在某种神秘的X射线之下,这种神秘的X射线持久而又不断地促使他们更不可能与社区保持联系。无论那种力量是什么,这种力量(而非发生于20世纪70和80年代的任何事情)还是解释了大部分的公民离散,

---

[16] 在20世纪70年代和80年代所进行的调查中,几乎不包括生于19世纪晚期的调查对象,以至于我们无法可靠地辨别连续出生的年龄群组中的差异。但是,那些不充分的数据(并未出现在图1中)显示,世纪之交可能曾是公民参与上升的一个时期。同样,1970年后出生的调查对象几乎也未出现在全国调查中,以至于我们很难确定他们独特的世代性概貌,尽管截至目前有微弱的结果表明,在公民参与方面,40年一代的下降可能正停止下跌并即将回升。然而,即使这一点被证明是正确的,上述发展要抑制公民参与的总体下降,可能还需要几十年的时间,原因下文给出。

[17] 1910—1940年间的这一代人的成员比他们的上一辈似乎也更具公民精神,至少根据出现在我们样本中的生于19世纪晚期的相对较少的那部分男女的观点来判断,情况如此。

[18] 在1964—1976年间所进行的全国选举调查中,与社会信任相关的问题每两年就会出现,随后又在1992年的调查中再次出现。在分析过程中,我将1992年的NES的访谈也包括进来,目的就是要获得关于这样一些同一年龄段群组的估计数据,在早期所进行的调查中,这些人当时还太年轻,并未被包括进去。

这些公民离散是我们谜团的核心部分。……

重述这个谜

当代美国的公民不参与很大程度上具有世代性，这样说只是重新阐述了我们的主要谜题。我们现在知道，我们独自打保龄球的大部分原因，可能要追溯至20世纪40和50年代，而不是60和70年代。那种神秘的反公民的"X射线"是什么？它已影响到二战后成年的美国人，至少在进入70年代以后，其影响还在不断地加深。[19] ……

我已经注意到唯一的一个明显的可疑对象，我准备动用间接证据向它发起攻击。……罪魁祸首就是电视。

首先，向它发起攻击的时机已到。漫长的公民一代，是最后一代并未伴随电视而成长起来的美国公民，因为电视与照明一样，在20世纪50年代才突然进入美国社会。在1950年，勉强有10%的美国家庭拥有电视机，但到了1959年，90%的家庭都有电视机，这或许是先前所记录的最迅猛的技术革新扩散。这种雷光电掣般的扩散，所造成的余波持续了几十年，比如，整个20世纪60年代，人均看电视的时间增长17%～20%，而整个70年代又在这个基础上增长了7%～8%。早些年，看电视主要集中于受教育程度较低的群体，但在整个20世纪70年代，受教育程度较高的群体看电视的时间也开始呈上升趋势。看电视的时间随年龄（尤其是退休以后）而增长，但是，由于电视的引进，每一代人都在一个较高的起点上开始其生命周期。截止到1995年，每一电视用户看电视的时间，比其在20世纪50年代要高出50%多。[20]

---

[19] GSS对公民参与指标的其他分析（本文并未详细报告）证实了20世纪80年代中的这种向下的转变。

[20] 我在此记录的一个学说，要分别归功于Robert Salisbury（1985），Gerald Gamm，and Simon & Garfunkel。全民娱乐的爱好者可能还记得，迪马乔（Joe Dimaggio）1936年与洋基队（Yankees）签约，当时，漫长公民一代的最后一批人才刚刚参与棒球运动，而他于1951年将中外场交给曼托（Mickey Mantle）负责，这个时候恰恰又是最后的"乳儿"达到法定成年年龄的时候。几乎同时性地，Braves，Athletics，Browns，Senators，Dodgers，以及Giants等队都放弃了自己的主场城市，这些城市自19世纪晚期以来就一直是它们的主场。当曼托于1968年离开洋基队时，对公民忠诚造成极大伤害。这种解释说明了，为什么罗宾逊夫人的悲伤的疑问[与乔廷·乔（Joltin Joe）那年的下落有关]，会引起如此广泛的共鸣。关于社会资本下降的一种解构主义的分析，突出了最后难以忘怀的哀叹——"我们的国家用孤独无助的眼神望着你"。

大量研究估计,普通美国人如今每天看电视的时间大约是 4 个小时。[21] 罗宾逊(Robinson,1990b)运用比较保守的时间耗用技术(用以确定人们如何分配他们的时间),提供的一个估计数字是每天接近 3 小时;但他承认,作为一种主要活动,看电视占用普通美国人 40% 的自由时间,这一比例自 1965 年以来又大约上涨 1/3。而且,复合设备已经蔓延:截止到 20 世纪 80 年代末,3/4 的全美家庭都拥有不止一套的电视设备(Comstock,1989),而这个数字还在稳定上升。总之,正如罗宾逊和戈德贝(Robinson & Godbey,1995)所断言的那样,"电视威力无比,它侵占了人们大量的空闲时间"。美国人度过白昼和黑夜的方式已发生大规模变化,这种变化正好出现在世代性的公民不参与的那些年中。

然而,电视来临和社会联系削弱发生关联的证据并不仅仅是间接性的。公民参与和电视收看之间的关系可以同公民参与和报纸阅读之间的关系相比较,这种比较具有启发意义。基本的对比简单明了:报纸阅读与高数量的社会资本相关,电视收看与低数量的社会资本相关。

限定教育、收入、年龄、种族、住所、工作状况以及性别等变量,那么,电视收看与社会资本及团体会员数之间具有很强的负相关性,而社会资本及团体会员数与报纸阅读之间则具有正相关性。……

电视如何破坏社会资本?

- 时间置换。尽管每个人的一天中都只有 24 小时,但多数形式的社会和媒介参与依然是正相关的。聆听大量古典音乐的人,比其他人更有可能(并非更不可能)参加 Cubs 比赛。电视是上述概括的一个首要例外——它是唯一的一种可能会禁止户外参与的休闲活动。电视收看所导致的结果,往往是牺牲几乎所有的户外社会活动,尤其是社会集会和非正式交谈(Comstock et al. 1978; Comstock 1989; Bower 1985; and Robinson & Godbey 1995)。电视观众就是有恋家癖的人。……
- 在加拿大的三个城市所进行的一项重要的与电视引进有关的准实验研究(Williams 1986),在跨时间的总体水平上发现了同样的模型:电视出现的一个主要影响就是导致社会、娱乐及社区活动的参与在所有年龄段的人群中都呈下降趋势。简言之,电视正将我们的闲暇时间私有化。
- 对观众视野的影响。一个令人印象深刻的文学团体(在"小人世界效应"的指导下聚集在一起)认为,过量观看电视的人,异乎寻常地怀疑他人的善行——例如,过高估计犯罪率。这个文学团体就深层的因果模式产生了大量争论,其中一些怀疑论者认为,上述那种遁世心理可能助长了电视迷的行

---

[21] 有关电视社会学方面的大量文献的介绍,可参见 Bower (1985), Comstock et al. (1978), Comstock (1989), and Grabner (1993)。本文出现的收视时间方面的数据,均引自 Bower (1985, 33)以及《公共视点》(*Public Perspective*,1995, 47)这本书。年龄群组的差异也引自 Bower (1985, 46)的报告。

为,反之不然。然而,在等待更好的实验证据的同时,我们还是可以接受一个合理的暂行判断:过量的电视收看很可能会增加关于人性的悲观想法(Gerbner et al., 1980; Dobb & MacDonald 1979; Hirsh 1980; and Comstock 1989, 265-69)。……

- 对孩子们的影响。电视在孩子们的生活中充当特别的角色——他们平均每周要耗费40小时的时间来看电视。收视率在青春期以前的孩子们当中尤其高,但它在青少年中依然很高:时间耗用研究(卡内基青少年发展委员会,1993, 5,引自 Timmer et al., 1985)表明,在9—14岁的青少年中,电视耗费的时间相当于所有其他自由活动(包括玩耍、业余爱好、俱乐部活动、户外活动、非正式拜访,以及闲逛)所耗费时间的总和。当然,电视对孩子社会化的影响已经被热烈讨论了30多年。从一些杂乱无章的有时相互矛盾的结果中,可得出的一个最合理的结论似乎就是:过量的电视收看可能会提高攻击性(尽管可能不是实际暴力),还可能会降低学业成绩,而且它在统计上与“社会心理障碍”有关,尽管这种影响的程度具有自我选择性,并且因果关联的程度尚有大量争议(Condry, 1993)。如我所说,这些证据尚不足以将其(电视)定罪,但对这些证据的辩护已作了大量说明。

结　论

伊契尔·索勒·普尔(Ithiel de Sola Pool)死后出版的《无边界的技术》(*Technologies Without Borders*,1990),是一本有先见的著作,书中的预见性观点,与我们目前关于科技、公共政策和文化这三者间的复杂联系的全国性争论,惊人地相关。普尔捍卫一种他所说的“温和的技术决定论”。通讯技术革命已经对社会生活和文化产生了深远影响,比如,印刷机有助于促进技术革新。普尔断定,通讯技术中的电子革命(甚至在我们意识到迫在眉睫的变化之前,他早就勾勒出了电子革命的概貌),是几个世纪以来最重要的技术进步,它将对社会和文化产生一种深远的分散性及分裂性的影响。

普尔所希望的结果可能是“没有邻接的社区”。作为一名杰出的自由主义者,他热忱欢迎技术变革为个体自由带来的好处,在某种程度上,我和他一样拥有那份热情。我们当中为当代美国社区的减少而感到惋惜的那些人,有必要对自由主义者在同一时期所取得的进步作出回应。我们必须避免一种不假批判地向往50年代的怀旧之情。另一方面,普尔又预言,同样一些与自由“友好”的技术,其出现实际上是暗中破坏我们之间的相互联系,以及我们同我们社区之间的联系。我想,普尔也会接受上述论证,因为普尔最有才干的一位门生,萨缪尔·波波金(Samuel Popkin,1991, 226-231)已经论证,电视的出现以及社会互动的相应下降,已经削弱了美国的政治话语权。普尔生前的最后一本书(Pool,1990, 262)的最后一行这样写道:“我们可以料想,(我们能够预料的技术发展趋势)将促进个人主义,并且将

使我们更难(而不是更容易)统治和组织一个融洽的社会。"

普尔的技术决定论之所以是"温和的",正是因为他认识到,社会价值可以限定技术的影响。最终,这一观点促使我们并不仅仅要思考,技术正如何使我们的生活私有化(如果它是这样的话,在我看来它似乎是的);而是要问:我们是否完全愿意接受这个结果,如果不愿意,我们能够对它做些什么。但那是将来要讨论的话题。

参考文献

Bower, Robert T. 1985. *The Changing Television Audience in America.* New York: Columbia University Press.

Brehm, John, and Wendy Rahn. 1995. "*An Audit of the Deficit in Social Capital.*" Durham, NC: Duke University. Unpublished manuscript.

Butler, David, and Donald Stokes. 1974. *Political Change in Britain: The Evolution of Electoral Choice*, 2nd *ed.* New York: St. Martin's.

Caplow, Theodore, Howard M. Bahr, John Modell, and Bruce A. Chadwick. 1991. *Recent Social Trends in the United States*: 1960-1990. Montreal: McGill-Queen's University Press.

Carnegie Council on Adolescent Development. 1993. *A Matter of Time: Risk and Opportunity in the Nonschool Hours: Executive Summary.* New York: Carnegie Corporation of New York.

Coleman, James. 1990. *Foundations of Social Theory. Cambridge*, MA: Harvard University Press.

Comstock, George. 1989. *The Evolution of American Television.* Newbury Park, CA: Sage.

Comstock, George, Steven Chaffee, Natan Katzman, Maxwell McCombs, and Donald Roberts. 1978. *Television and Human Behavior.* New York: Columbia University Press.

Condry, John. 1993. "Thief of Time, Unfaithful Servant: Television and the American Child," *Daedalus* 122 (Winter): 259-78.

Converse, Philip E. 1976. *The Dynamics of Party Support: Cohort-Analyzing Party Identification.* Beverly Hills, CA: Sage.

Cutler, Blaine. 1990. "Where Does the Free Time Go?" *American Demographics* (November): 36-39.

Davis, James Allan, and Tom W. Smith. *General Social Surveys*, 1972-1994. [machine readable data file]. Principal Investigator, James A. Davis; Director and Co-Principal Investigator, Tom W. Smith. NORC ed. Chicago: National Opinion Research Center, producer, 1994; Storrs, CT: The Roper Center for Public Opinion Research, University of Connecticut distributor.

Dobb, Anthony N., and Glenn F. Macdonald. 1979. "Television Viewing and Fear of Victimization: Is the Relationship Causal?" *Journal of Personality and Social Psychology* 37: 170-79.

Edwards, Patricia Klobus, John N. Edwards, and Ann DeWitt Watts, "Women, Work, and Social Participation." *Journal of Voluntary Action Research* 13 (January-March, 1984), 7-22.

Elazar, Daniel J. 1966. *American Federalism: A View from the States.* New York: Crowell.

Fukuyama, Francis. 1995. *Trust: Social Virtues and the Creation of Property.* New York: The Free Press.

Gerbner, George, Larry Gross, Michael Morgan, and Nancy Signorielli. 1980. "The 'Mainstreaming' of America: Violence Profile No. 11," *Journal of Communication* 30 (Summer): 10-29.

Ginzberg, Eli. *The Unemployed.* 1943. New York: Harper and Brothers.

Glenn, Norval D. 1987. "Social Trends in the United States: Evidence from Sample Surveys." *Public Opinion Quarterly* 51: S109-S126.

Grabner, Doris A. 1993. *Mass Media and American Politics.* Washington, D. C.: CQ Press.

Hirsch, Paul M. "The 'Scary World' of the Nonviewer and Other Anomalies: A Reanalysis of Gerbner et al.'s Findings on Cultivation Analysis, Part I," *Communication Research* 7 (October): 403-56.

Hughes, Michael. 1980. "The Fruits of Cultivation Analysis: A Re-examination of the Effects of Television Watching on Fear of Victimization, Alienation, and the Approval of Violence." *Public Opinion Quarterly* 44: 287-303.

Jahoda, Marie, Paul Lazarsfeld, and Hans Zeisel. 1933. *Marienthal.* Chicago: Aldine-Atherton.

Mannheim, Karl. 1952. "The Problems of Generations." In *Essays on the Sociology of Knowledge*, Ed. Paul Kecskemeti. New York: Oxford University Press: 276-322.

Meyrowitz, Joshua. 1985. *No Sense of Place: The Impact of Electronic Media on Social Behavior.* New York: Oxford University Press.

Miller, Warren E. 1992. "The Puzzle Transformed: Explaining Declining Turnout." *Political Behavior* 14: 1-43.

Miller, Warren F., and J. Merrill Shanks. 1995. *The American Voter Reconsidered.* Tempe, AZ: Arizona State University. Unpublished Manuscript.

Pool, Ithiel de Sola. 1973. "Public Opinion." In *Handbook of Communication*, ed. Ithiel de Sola Pool et al. Chicago: Rand McNally: 779-835.

Pool, Ithiel de Sola. 1990. *Technologies Without Boundaries: On Telecommunication in a Global Age.* Cambridge, MA: Harvard University Press.

Popkin, Samuel L. 1991. *The Reasoning Voter.* Chicago: University of Chicago Press.

Postman, Neil. 1985. *Amusing Ourselves to Death: Public Discourse in the Age of Show Business.* New York: Viking-Penguin Books.

Public Perspectives. 1995. "People, Opinion, and Polls: American Popular Culture." 6 (August/September): 37-48.

Putnam, Robert D. 1993. *Making Democracy Work: Civic Traditions in Modern Italy. Princeton*, NJ: Princeton University Press.

Putnam, Robert D. 1995. "Bowling Alone, Revisited," *The Responsive Community* (Spring): 18-33.

Putnam, Robert D. 1996. "*Bowling Alone: Democracy in America at the End of Twentieth Century*," forthcoming in a collective volume edited by Axel Hadenius. New York: Cambridge University Press.

Robinson, John. 1981. "Television and Leisure Time: A New Scenario," *Journal of Communication* 31 (Winter): 120-130.

Robinson, John. 1990a. "The Time Squeeze," *American Demographics* (February).

Robinson, John. 1990b. "I Love My TV." *American Demographics* (September): 24-27.

Robinson, John, and Geoffrey Godbey. 1995. *Time for Life.* College Park, MD: University of Maryland. Unpublished Manuscript.

Rosenstone, Steven J., and John Mark Hansen. 1993. *Mobilization, Participation, and Democracy in America.* New York: Macmillan.

Salisbury, Robert H. 1985. "Blame Dismal World Conditions on ... Baseball." *Miami Herald* (May 18): 27A.

Schor, Juliet. 1991. *The Overworked American.* New York: Basic Books.

Sharkansky, Ira. 1969. "The Utility of Elazar's Political Culture." *Polity* 2: 66-83.

The Economist. 1995. "The Solitary Bowler." 334 (18 February): 21-22.

Timmer, S. G., J. Eccles, and I. O'Brien. 1985. "How Children Use Time." In *Time, Goods, and Well-Being*, ed. F. T. Juster and F. B. Stafford. Ann Arbor, MI: University of Michigan, Institute for Social Research.

U. S. Bureau of the Census. 1995 (and earlier years). *Current Population Reports.* Washington, D. C.

Verba, Sidney, Kay Lehman Schlozman, and Henry E. Brady. 1995. *Voice and Equality: Civic Volunteerism in American Politics.* Cambridge, MA: Harvard University Press.

Wilcock, Richard, and Walter H. Franke. 1963. *Unwanted Workers.* New York: Free Press of Glencoe.

Williams, Tannis Macbeth, ed. 1986. *The Impact of Television: A Natural Experiment in Three Communities.* New York: Academic Press.

关于作者

罗伯特·D. 普特南(Robert D. Putnam)是哈佛大学"国际事务研究中心"的教授和主管。其1993年的著作,《使民主运作起来:现代意大利的公民传统》(*Making Democracy Work: Civic Traditions in Modern Italy*)(普林斯顿大学出版社),获得APSA"比较政治学组织分部"的1994年的Gregory Luebbert奖。

累积和限制投票制下的少数族群的代表性[①]

附录 B

"Minority Representation under Cumulative and Limited Voting"

大卫·布洛金顿(David Brockington)/华盛顿大学

托德·多诺万(Todd Donovan)/西华盛顿大学

肖恩·鲍勒(Shaun Bowler)/加利福尼亚大学里弗德赛分校

罗伯特·布里斯切托(Robert Brischetto)/德克萨斯州大学圣·安东尼奥分校

我们研究了20世纪90年代美国地方采取经修正的不分区选举制度(累积投票制与限制投票制)所导致的少数族群的代表性。我们提出并检验了关于不同地方选举制度下的描述性代表的相对比例的假说。我们发现,CV/LV选举所产生的描述性的非洲裔美国人代表,在水平上与那些在更大的单一席位选区所产生的代表性接近,并超过那些来自一些小的南部单一席位选区的代表性水平。关于拉美人代表性的研究结果更具说服力。我们的结果使一些人欢欣鼓舞,这些人对推动少数族群的代表性感兴趣,而这种推动又不必运用在种族基础上划分选区的苛刻程序。

为了少数族群的代表性,美国最高法院的近期决策已将更多的注意力指向选区划分的各种备选方案(Pildes & Donoghue,1995)。在本文中,我们对修正后的不分区选举方案(限制投票制和累积投票制)如何能够推进少数族群的代表性进行评估。我们还将考察,如何将这些方案下的代表性与划分选区和不分区选举制下所产生的代表性进行比较。

在前几十年,法院关于《投票权利法案》(VRA)的解释,扩大了司法部和少数族群起诉人的能力,使他们能够对那些可能削弱少数族群选票分量的地方选举方案提出挑战(例如,*Thornburg v. Gingles*,1986;*Gomez v. Watsonville*,1988)。通常最易受到挑战的方案包括,多席位不分区选举制度下选举产生的议会。对这些情况

---

① *The Journal of Politics*, Vol. 60, No. 4, November 1998, pp. 1108-1125

的常规补救措施，是调整为单一席位选区（SMD）方案，并且选区划分有助于促进少数族群的代表性。事实上，大量实质性的证据表明，种族及民族性的少数族群更有可能赢得与他们在分区辖区内的人口份额成比例的席位（Engstrom & McDonald，1981；Polinard，Wrinkle，and Longoria，1991；Welch，1990）。

对选区划分的限制：作为一种补救措施

然而，以种族为基础划分选区已越来越多地受到法院的仔细审查。“肖诉雷诺”案（1993）批评了“奇形怪状”的选区。“米勒诉约翰逊案”（1995）发现，多数—少数国会选区是违反宪法的，并论证说，选区不应当“从根本上以种族为基础进行划分，而不顾常规的及传统的选区划分实践”。“肖诉亨特”案（1996）以及“布什诉维拉”案（1996）发现，个别选区划分方案违反了《第十四修正案》的平等保护条例。一旦发现种族是选区划分的“主导因素”，当地法院必须对平等保护条例使用严格的审查标准。这使得各州及地方政府很难对采纳这样的选区产生真正的兴趣。

多数—少数选区同样受到指责，因为它们可能“浪费”选票（Lijphart，1994；Still，1984）。还有人认为，划分选区可限制少数族群对政策的影响（Guinier，1991，1994；Sass & Mehay n. d.），并阻止跨越种族界限的政党联盟的形成（Swain，1993）。这些建议还是存在实践上的问题。爱德华·斯蒂尔（Edward Still，1991）注意到，按非洲裔美国人占其总人口65%的代表比例划分选区，或许是在某些情况下促进非洲裔美国人代表性所要求的不能再低的最低限，尽管布雷斯等人（Brace et al.，1988）发现，这个最低比例因地而异。

美国的累积和限制投票制

为了应对已察觉到的选区划分的局限性，有人提议，可采用累积投票制（CV）和限制投票制（LV）来提高少数族群的代表性。大量中小型的美国辖区已经采纳了这些方案（Amy 1993；Cole & Taebel，1992；Cole，Taebel，and Engstrom，1990；Still，1984，1991）。② 两种制度的采用都是为了选举代表全州的多席位议会，并通过对选民投票方式的改变而推进代表比例的均衡。

北卡罗莱纳州、阿拉巴马州和得克萨斯州的一些辖区，已经采纳修正后的不分区选举制度（新墨西哥州、南达科塔州和伊利诺斯州的一些地方也适当采纳了这种制度）。在我们写作这篇文章时，超过75%的市议会、郡议会和学校董事会，都已

② 在美国，只有一个人口超过100 000的辖区采用CV或LV制度：伊利诺伊州的皮奥里亚。

采纳 CV 或 LV。[3] 这些选举产生相对较低的排斥临界值，即某一团体在假定所有团体的投票者都支持某一候选人的情况下，为了选举这个候选人所需的选票比例。例如，在五席位议会的 CV 选举中，某一少数族群候选人即使只获得 17% 的支持率，也不能拒绝其参加对席位的竞争。

在限制投票制下，投票者受到限制，他们拥有的选票要比有待选举的席位少。候选人当选遵循相对多数票原则，获得票数最多的当选，直到所有席位都被填满。在政党制度中，有人认为，LV 下的结果越具有比例性，与有待选举的席位数量相关的选票就越受限制（相关说明请参见 Lakeman，1970，80-88；Still，1984，253-55）。累积投票制对不分区选举制的修正方法是：选民的可投票数与应选席位数相同，另外，还可以在任何候选人组合之间分配聚集起来的选票。选民可按照自己的意愿任意分配选票，即使所投候选人数量少于应选席位数。候选人当选遵循相对多数原则，获得票数最多者当选，直到所有席位都被填满……

限制和累积投票制下的策略负担

有理由认为，实行限制投票制和累积投票制的地方，成比例的描述性代表会出现一些偏差。在选举制度分类中，CV 和 LV 通常被贴上“半比例性”的标签（Amy 1993；Lakeman 1970）。这些修正后的不分区选举方案之所以产生偏离比例性的偏差，大部分潜在原因就在于，对每一种制度在大众层面以及精英层面上所制定的战略合作（Cox 1997）的需求。在每一种制度下，某一政党或一个获得提名的团体，必须通过控制候选人选举进程的方式（比如，他们列举大量最理想的候选人名单），有效地将席位最大化。为了使代表最优化，这些团体还必须将支持者的选票准确地分布在他们所提名的那些候选人身上（Still，1984，254-255）。……如果一个团体提名过多并列举太多的候选人，那么，它就要承担因支持者的选票过于分散而支持率不足的风险，这会导致代表名额的不足。如果一个团体所提名的候选人不足，可能会导致浪费选票的错误，他们所浪费的选票也许会产生另外一个席位。

……为了确保一定的代表性，一个少数族群团体通常只需提名一个候选人（或者干脆在选票允许的范围内，尽可能多地提名候选人）。……在累积投票制下，对提名名单的控制对少数族群的一定比例的描述性代表而言是必需的，但是，累积投票制进一步修正了不分区选举方案，以至于可能会给少数族群造成额外的策略负担。运用累积投票制，参选团体必须协调他们的支持者的投票行为。换言之，投票者必须被告知在候选人之间分配选票的最优策略。无疑，无论是在大众层面（选票

③ 每个辖区的具体制度细节，可通过与各地地方官员的电话沟通来鉴定。与地方官员以及卷入 VRA 诉讼的个人之间的交流能够使我们鉴定这些社区。另外一些地方正着手解决能够导致采纳限制投票制或累积投票制的诉讼案。

分配)还是在精英层面(候选人提名)上,CV 都要求策略调和,而 LV 可能需要较少的策略调和,倘使投票仅限于一个候选人的话。因此,斯蒂尔(Still,1984, 256)指出,为了获得比例均衡的结果,CV 比 LV 可能需要更多的策略性投票行为(基于少数族群投票者在 CV 中犯策略错误的可能性;还可参见 Engstrom 1993)。

由于 CV 允许投票者在投票过程中拥有更大的选择权,所以,即使只有一个少数族群的候选人参加选举,这些机会的存在还是提高了一些少数族群的选民在多个候选人之间分散选票的概率。这相应地又增加了少数族群选票可能流向非少数族群候选人的几率,这潜在地限制了少数族群的有效选票转化为(应得)席位。[④]而且,在实行 CV 或 LV 的地方,如果某团体的人口份额接近排斥临界值,并且投票行为具有种族上的两极化,那么在这种情况下,少数族群的候选人如果要当选,唯一的可能就是他们的支持者积极出来投票,投票人数的比例能与多数团体的选民相抗衡。所有这一切都表明,相比在 LV 之下所获得的代表性而言,CV 下的描述性少数族群代表,可能在一定程度上同总人口更不成比例。

有理由认为,相比单一席位选区制(SMD),每一种方案都可能产生不怎么成比例的描述性代表。采用 SMD 方案,如果谋划者有能力裁定选区边界,以便大量地创建少数族群选区,并创造许多在一个辖区内与整个多数派选票份额相当的少数族群选区,那么,产生成比例的描述性代表就有指望。与 CV 和 LV 相比,一旦均质的多数派—少数派选区被创造出来,那么,为了产生比例均衡的描述性代表,对精英策略行为的要求就很有限(例如,动员大量少数族群的投票者设法接近多数派团体的投票者,并控制提名名单),并且对选民策略性投票行为的要求也很少(例如,散布选票)。[⑤] 换句话说,选区划分能够使推进少数族群代表性所需要的一些策略行为制度化,而且可能比 CV 和 LV 产生更多的少数族群代表。

假说:修正后的不分区投票制的结果

上述讨论间接表明了一些可检验的主张,这些主张与以下问题有关:修正后的不分区(替代方案)选举中的席位—人口关系,如何与那些在分区选举以及未修正的不分区(AL)选举方案下所获得的关系进行比较,以及 LV 和 CV 替代方案相互之间如何进行比较。

---

④ 对一个运用 CV 的城市的调查报告显示,36% 的拉美裔选民利用可选择的权利将他们的至少一张选票投向非拉美裔的候选人(Cole, Taebel, and Engstrom,1990; Cole & Taebel, 1992)。

⑤ 当少数族群团体在空间上高度隔离的时候,最优的选区划分安排便成为可能。由于非洲裔美国人的居住地一直以来都比拉美人的居住地更脱离白人的居住地(Massey & Denton, 1987),所以,相比拉美人而言,选区划分可能会产生更成比例的非洲裔美国人的代表(Taebel, 1978; 另参见 Vedlitz & Johnson,1982)。

首先，我们认为，修正后的不分区选举制度（CV 和 LV）应当比传统的不分区方法产生更成比例的少数族群代表。假定某团体大致结成一伙进行投票，那么，赢得相对多数的任何一个团体，均有可能在未修正的不分区选举中将所有席位尽收囊中（Lakeman，1970）。尽管几乎所有投票制度都偏向选举过程中获得最多选票份额的团体，但在美国风格的不分区的、相对多数票制下，这种偏向最严重（Johnson，1979）。相反，LV 和 CV 均比不分区选举制具有更低的排斥临界值……

其次，一旦策略要求和 CV/LV 的"半比例性"的本质已定，我们便能料想到，相比 SMD 选举而言，修正后的不分区选举可能会产生相对不成比例的描述性的少数族群代表。[6] 修正后的不分区选举方案，降低了某一少数族群候选人赢得席位所需选票的最低限，通过这种方式，它可以推进少数族群的代表性，但是，参选团体必须在政治上很好地组织起来，才能掌控这些制度。单一席位选区制对精英和大众的合作要求相对较低，在此情况下，便能推进少数族群的代表性。

再者，限制投票制比累积投票制有望产生更成比例的结果。这一假设建立在如下假定的基础之上：累积投票制可能要求投票者和精英之间某种额外的策略行为因素（调和分散的选票）。

数据及分析框架

我们所检验的案例均取自美国的市、郡以及学校的选区，这些选区都采纳 CV 或 LV，以应对实际的或预期的 VRA 诉讼。相对于少数派团体在当地总人口中所占的比重而言，选举制度如何与他们的代表性相关，由于我们对研究这一问题感兴趣，所以我们只分析 1990 年人口普查所覆盖的一些辖区。有关地方选举制度和选举结果的原始信息，是在 1995 年春通过与市政秘书及乡镇选举官员的电话访谈获得的，随后在 1996 年与 1997 年又进行了访谈，获得了更多数据。我们所验证的几乎所有地方都位于这三个州：得克萨斯州、阿拉巴马州和北卡罗莱纳州。南达科塔州、新墨西哥州和伊利诺斯州也各有一个辖区采用了累积投票制。我们将我们的分析限定在来自以上六个州的一些地方，在这些地方，占主导地位的少数族群总共也不及投票人口的 50%。

我们将单次选举作为分析案例。这增加了分析的可比性，主要原因在于，这种做法减弱了那些将一些选举交错进行的辖区所引起的问题。我们的分析包含每一辖区的两次最近的选举；如果某辖区截止到 1997 年春只进行一次 CV 或 LV 竞选，则选取最近的一次进行分析。这使得我们能够捕捉地区间及地区内各项选举间的

---

[6] 这种逻辑并非意味着，相比多席位的 LV 或 CV 制度而言，任何一个（全国性的）运用"赢者通吃"、单一席位选区的选举制度，就一定会产生出更成比例的结果。注意到这一点很重要：我们所料想的比例差异，与其说是单一席位选区本身的一个函数，不如说是种族性和民族性群体被指派到特定选区的一个函数。

表 1　个案的种族/民族分布

占主导地位的少数派团体	总选举次数	少数族群候选人参加的选举次数	少数族群获胜的选举次数
美籍拉美人	66	47(71%)[a]	33(50%)[a] (70%)[b]
非洲裔美国人	28	26(93%)[a]	25(89%)[a] (96%)[b]
美国土著人	02	01(50%)[a]	00(0%)[a b]

a. 所有选举次数的百分比；

b. 有少数族群候选人参加的选举次数百分比；

注：单次选举作为案例。数据来自最近两次选举；如果某一辖区截止到 1997 年只进行过一次累积投票制或限制投票制竞选，则数据来源于最近举行的一次选举。

差异，因为大多数社区在连续选举中均改变了他们所角逐的席位数量。我们的数据来源中，只有三个地方的倒数第二次选举是在 1990 至 1992 年间举行的，其他所有地方的最近两次竞选都是在 1994 至 1997 年间进行的。

因变量是少数族群候选人在每次选举中所赢得的 CV 或 LV 席位的百分比。这样一来，我们的模型就将席位—人口关系与修正后的不分区选举之间的独特关系分离出来，并在分析中排除了这些团体通过其他选举方法所赢得的席位。⑦ 席位—人口分析所涉及的社区仅限于人口超过 1 000 的地方，这是人口普查数据中包含种族信息的最低人口数量。大多数社区人口规模相当小，中位数为 3 167 人，平均数为 10 311 人。用以评估席位—人口关系的最终样本，包括 62 个辖区所提供的 96 次选举的全部数据。

表 1 表明，少数族群在角逐这些选举时曾取得过成功，尽管在选举制度改变以前，其代表性程度处在较低的水平上。在累积投票制下，在拉美候选人谋得官职的 70% 的选举角逐中，都有拉美候选人参选。而且，在非洲裔美国人谋得官职的 96% 的 CV/LV 选举中，至少一位非洲裔美国人当选。表 1 还具体说明了我们将在下文讨论的内容：占主导地位的少数族群为拉美人的选举中，仅有 71% 的选举有拉丁裔候选人参加。

关于如何将少数族群达到投票年龄的人口份额转化为地方议会中的席位的假说，可以通过对相对于人口的席位进行回归分析的方式来检验（Engstrom & McDonald，1981）。双变量回归产生斜率估计值，它们可以被用来估计不同选举制度下的席位—人口关系的差异，并且可以与其他使用类似方法的研究所产生的结果相比较。例如，当运用百分比数据时，截距为 0 的 1.0 的斜率表明，少数族群在

---

⑦ 例如，倘若某地举行的选举中，有五个席位是通过累积投票制（或限制投票制）产生的，则因变量的计算方法为：少数族群候选人赢得的累积投票制产生的席位数/5。类似地，如果五个席位中有三个是通过划分选区方法产生的、两个是通过累积投票制产生的，那么因变量的计算方法则为：少数族群候选人赢得的累积投票制产生的席位数/2。

地方议会中所占的席位份额，与当地达到投票年龄的少数族群人口的百分数始终保持一致……

如表 2 所示，用于所有 CV 和 LV 辖区的模型首次得到评估，案例中占主导地位的少数派团体，包括非洲裔美国人、美籍拉美人和美国土著人（模型 2.1）。[⑧] 接下来，我们评估了分别对应于非洲裔美国地区和拉美地区的两组模型，旨在评定，对这些团体而言，修正后的不分区投票制是否与一种不同的席位—人口关系相关（分别为模型 2.2 和 2.3）。而且，我们还运用交互作用项和虚拟变量，评估用于所有案例的多重变量模型，以检验少数族群的代表性在 LV 下是否比在 CV 下更接近常规标准（模型 2.4）。为了评价其他假说，我们将这些模型所产生的斜率，与其他研究所产生的斜率进行比较。这种比较可为下述问题提供一些看法：经修正的不分区选举制下的少数族群的代表性，如何与美国地方所采用的其他选举制度进行比较。

有关这些关系的斜率的信息比一种简单的证明更重要，这种简单证明即是：在调整为经修正的不分区选举制以后，越来越多的少数族群加入地方议会。每一个辖区都采用了新的选举制度，因为它们都拥有数量可观的少数族群人口，而少数族群的代表相当有限（多数地方甚至没有少数族群的代表）。它们当中只有两个地

表 2　累积投票制与限制投票制下的少数族群的席位—人口关系

变量	模型 2.1（所有案例）	模型 2.2（黑人案例）	模型 2.3（拉美人案例）	模型 2.4（LV）
少数族群法定选举人口百分比	0.15	0.95***	0.03	0.01
	(0.18)	(0.13)	(0.21)	(0.18)
LV 虚拟变量	—	—	—	-0.19
				(0.15)
LV* 选举制下少数族群法定选举人口百分比	—	—	—	1.12*
				(0.48)
截距	0.15**	0.04	0.16	0.17**
	(0.06)	(0.08)	(0.07)	(0.06)
R^2	0.01	0.26	0.00	0.12[a]
案例数量	96	28	66	96

a. 模型 2.4 的 R^2 是调整过的；其他都未调整。

** $p < 0.05$；*** $p < 0.01$（双尾 t—检验）

注：因变量 = 每次选举中少数族群候选人赢得的累积投票制或限制投票制席位数/每次选举中总共的累积投票制或限制投票制席位数。表格中填写的是非标准回归系数，或者因变量与自变量之间关系的斜率。

⑧　在一个只实行累计投票制的辖区（Sisseton，SD 学区），占主导地位的少数派团体是美国土著人。

方的少数族群人口少于10%（最少的为7%）。[9] 我们的分析所涉及的多数（即使不是大部分）辖区，在修正后的不分区选举制下，少数族群的代表性得以提高。我们的模型说明了这一过程的系统性本质，在某种意义上推动了与其他选举制度的研究的比较。

结　果

模型2.1的结果显示，当所有经修正的不分区辖区同时被考察时，少数族群的人口和少数派群体所控制的席位百分比并不匹配。当我们分别考察非洲裔美国辖区和拉美辖区时，经修正的不分区选区之间的席位—人口关系的变化是明显的。当我们单独考察非洲裔美国辖区时（当分析局限于黑人候选人的案例时，R^2 增加至0.45），这一模型的拟合度大大提高（$R^2 = 0.26$）。模型2.2表明，相对于这些选区而言，席位和少数族群法定投票人口之间的关系，可由一条斜率为0.95、截距接近0（差异不显著）的斜线表示。因此，在主要少数族群为非洲裔美国人的地方，随着人口中少数族群份额的增加，描述性代表也相应地增加。而且，由于截距接近0，所以人口份额在较低的水平上转化为少数族群人口的代表。

针对占主导的少数族群为拉美人的地方，模型2.3指明了席位—人口关系的斜率。倘若斜率不显著并且 R^2 极低，那么，少数族群的人口份额和少数族群在议会及学校董事会中所占的席位份额之间，似乎就不存在实质的线性关系。这并非意味着，在采纳经修正的不分区选举制的地方，拉美候选人并未当选。与多数非洲裔美国辖区一样，在调整为经修正的不分区选举制以后，拉美候选人大量当选。如表1所示，在拉美候选人谋求公职的案例中，有超过一半（33/47）的情况是成功的——相比这些社区先前的拉美裔的代表性而言，这标志着一种巨大的改进（Brischetto，1995；Brischetto and Engstrom，1998）。然而，很多拥有大量拉美裔人口的地方，选举却没有产生拉美裔代表，究其原因，可归咎为拉美裔候选人的缺乏（以及）民意调查的缺陷。就算只分析拉美裔候选人谋求公职的案例，效果仍不显著（本文未作阐述）。在讨论部分，我们将对席位—人口模型如何低估拉美裔代表性的潜在性进行阐述。

---

[9] 例如，托马斯和斯特瓦特（Thomas and Stewart，1988，171）注意到，1982年联邦政府所研究的阿拉巴马州占主导地位的黑人区中，有44%的乡村没有黑人代表（美国公民权利委员会，1983，引用自 Thomas and Stewart，1988）。这些没有黑人代表的乡村中，大部分都是黑人人口占多数。许多城镇都是由这些乡村组成的。自1964年以来，北卡洛莱纳州中每一个采用LV的地区，都在《投票权利法案》的管辖范围之内。基奇和西斯特罗姆（Keech and Sistrom，1994）报告说，截止到1989年，90%的北卡洛莱纳州乡村和城市使用的是未修正的不分区选举制。这些地方的黑人代表名额严重不足。在黑人是少数族群的地方，代表权益分值不超过0.20。几乎所有的北卡洛莱纳州的案例都来自这些社区。

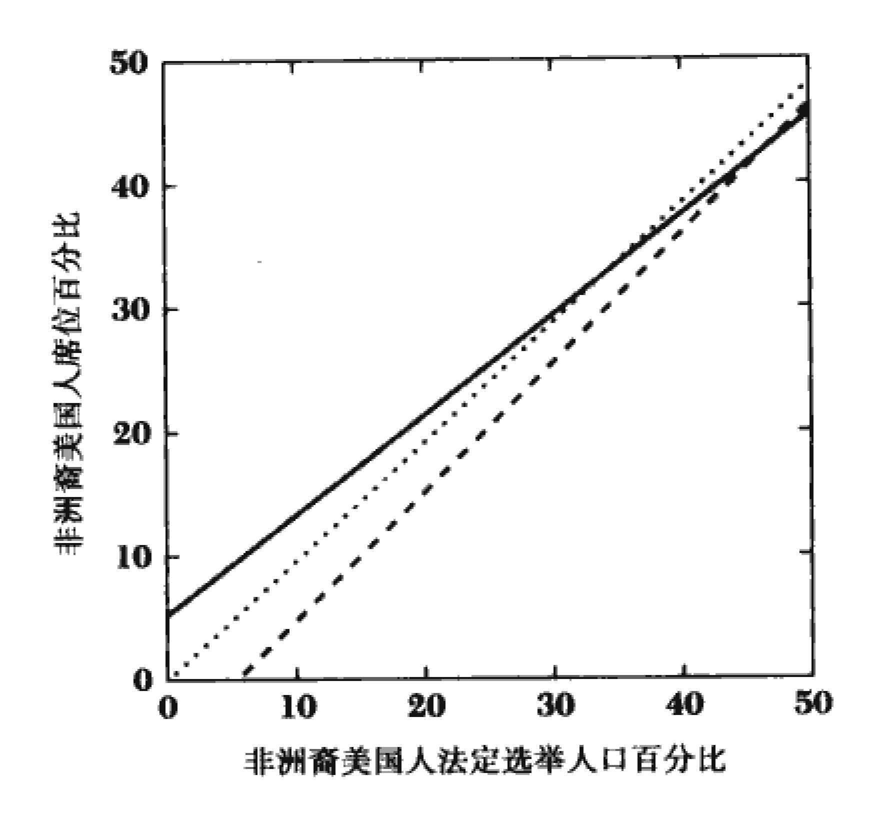


图1 CV/LV 席位—人口关系与美国南部其他选举制度的比较

实线 = 选区制(Welch,1990)

虚线 = 不分区制(Welch,1990)

点线 = 修正后的不分区选举制(CV/LV)

我们假设,相比标准的不分区选举方案而言,这些新的经修正的不分区制度将产生更成比例的描述性代表,而相比 SMD 方案而言,这些制度又产生出更不成比例的代表。将我们的模型的结果与其他的考察美国地方选举制度的研究结果进行比较,我们的结果被证明是合理的。如果我们将经修正的不分区制度下的非洲裔美国人代表性的斜率和截距,同关于标准的不分区选举制的先前研究的斜率和截距进行比较,我们会发现一些证据支持这些假说。

如图 1 所示,近期经修正的不分区投票制下的非洲裔美国人的席位—人口关系,与威尔奇(Welch,1990)所发现的少数族群为黑人的采用 SMD 的南方大城市(居民超过 5 万人)中的非洲裔美国人的席位—人口关系类似。图 1 将我们的模型 2.2 的结果与威尔奇的发现进行了比较。综观少数族群人口范围的许多区域,修正方案(CV/LV,由点线表示)所产生的代表水平似乎与分区制(实线)所产生的代表水平类似,并且比标准的不分区方案(虚线)产生出稍微多一些的描述性代表。

由于我们近乎所有的案例均来自相对较小的选区的选举,所以威尔奇的数据可能不是最好的适宜于比较跨选举制度的席位—人口关系的数据。另外有一项对其他最近放弃 AL 方案的南方选区的研究,这项研究包含许多与本研究所包含的社区更为类似的较小的乡村选区。当把我们的估计值与这项研究的结果相比较时,我们发现,在修正后的不分区选举方案下,非洲裔美国人的代表水平更为突出。布洛克(Bullock,1994)研究了 1991 年乔治亚州乡委员会选举情况。所研究的乡镇

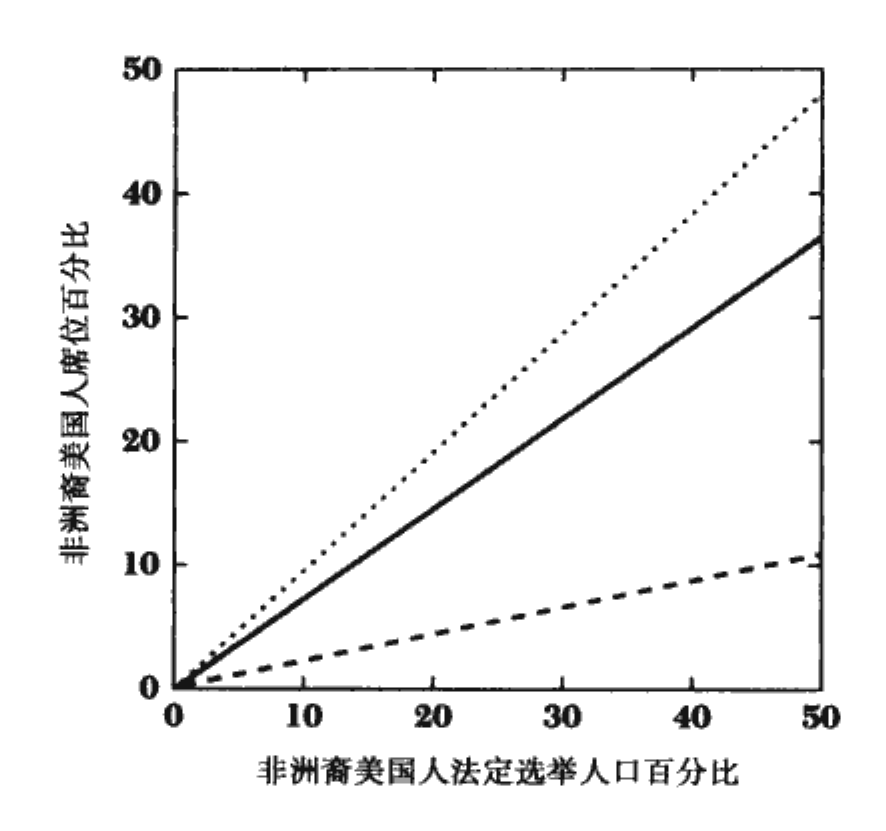


图 2　CV/LV 席位—人口关系与南部较小地区的其他选举制度的比较

实线 = 选区制(Bullock,1994)

虚线 = 不分区制(Bullock,1994)

点线 = 修正后的不分区选举制(CV/LV)

中,有 52 个采用单一席位选区制,比 1981 年的 17 个有所增长。[10] 图 2 绘制了单一席位选区制下的布洛克的席位—参加投票人数的双变量关系图(实线),以及我们在模型 2.2 中所评估的使用 CV/LV 方案的非洲裔美国辖区的关系图(点线)。该图所显示的结果与我们的一个假设相悖——当与分区选区相比时,经修正的不分区选举实际上会产生稍微更成比例的结果。

当我们将我们的估计值与布洛克关于乔治亚州乡镇的研究结果进行比较时,我们发现,标准的不分区选举与 CV/LV 选举之间存在实质差异,并且这种差异与我们的一个假说相容,这个假说就是:CV/LV 方案比传统的 AL 制度能产生更成比例的少数族群的代表。图中虚线所表示的是,1991 年仍然使用未修正的 AL 选举制的 41 个乔治亚州乡镇的双变量的席位—参加投票人数关系。与这些案例相比,经修正的不分区选举制实质上产生了更多的非洲裔美国人的描述性代表。[11]

我们的最后一个假说处理 LV 和 CV 制度下的结果间的差异。由于 LV 方案可

---

[10] 与我们当时所研究的案例类似,许多地方都刚刚放弃 AL 方案,以应对第二节所提到的实际的或潜在的违反《投票权利法案》的行为。与我们的研究一样,布洛克的研究案例也局限于少数族群为非洲裔美国人的地方($N=149$ 个乡镇)。由于大多数的这些选区都刚刚转换选举制度,而且大多数地方都具有人口稀少和乡村特点,我们认为它们是与本研究作比较的良好参照体。

[11] 我们的模型将少数族群的法定选举人口的百分比用作关键自变量,而布洛克所使用的自变量是少数族群投票人数的百分比,考虑到这一情况,我们对 CV/LV 的估计与布洛克对不分区选举制和单一席位选举制的估计之间的差异更明显。倘若人口与选民登记数之间存在差距,那么我们的预测便有可能偏离发现成比例的关系,当运用参加登记投票人数时,这种现象是明显的。

表 3 累积投票制与限制投票制下的非洲裔美国人的席位—人口关系

变量	模型 3.1（仅限累积投票制）	模型 3.2（仅限限制投票制）
少数族群法定选举人口百分比	0.60**	1.12*
	(0.25)	(0.61)
截距	0.10**	-0.02
	(0.04)	(0.19)
R^2	0.38	0.19
案例数量	11	17

*p < 0.10; **p < 0.05(双尾 t-检验)

注:因变量 = 每次选举中少数族群候选人赢得的 CV 或 LV 席位数/每次选举中总共的 CV 或 LV 席位数。表格中填写的是非标准回归系数。

能涉及较少的策略负担(例如选票分布调和),所以我们期望,限制投票制能够比累积投票制产生更多的少数族群代表。模型 2.4 包含一个系数,它反映了表示 LV 选区的虚设物和少数族群法定投票人口之间的交互作用。理论上,这一系数将累积投票制下的席位—人口关系单独分离出来。当我们比较 LV 和 CV 选区时,交互作用项的显著系数(1.12; p <0.05)可被认为反映了更成比例的少数族群的描述性代表。这与有效使用 LV 对策略要求更低的假设一致。

有必要强调:表 2(N = 17)的分析包含相对较少的 LV 案例,并且这些案例都来自占主导地位的少数族群为非洲裔美国人的地区。[12] 模型 2.4 中的交互作用,可能会反映出 LV 下的非洲裔美国人的代表性与 CV 下的拉美裔及非洲裔美国人的代表性之间的差异。

另外一种确定少数族群是否在 LV 下实现更高代表性的方法,就是评估 28 个含有非洲裔美国少数族群选区的席位—人口关系的斜率。如表 3 所示,我们首先评估使用累积投票制地区所独有的席位—人口关系的斜率(模型3.1),然后将其与使用限制投票制的非洲裔美国地区的席位—人口关系的斜率相比较(模型3.2)。表 3 记录了比较的结果。我们再次发现:有证据证明,LV 选举制(b =1.12, p = 0.08)下的少数族群人口与席位的关系,比累积投票制(b =0.60, p <0.05)下的少数族群人口与席位的关系更成比例(或者超过比例)。[13] 但是,LV 下的人口—席位系数并不明显大于 CV 下的系数。表 3 中的人口—席位模型的斜率和截距值的确显示,当黑人人口份额接近 50%时,CV 社区黑人候选人的当选率与布洛克所分析的分区选区的当选率几乎一样。在黑人人口较少的范围内,模型 3.1 的截距表明,CV 比布洛克的 SMD 选区产生更多的代表。

⑫ 不同的非洲裔美国地区之间,均会发生所有的选举变化。

⑬ 对这些斜率间差异的 t—检验并未产生显著的差异。小规模的样本限制了这一检验的效力。当把表 3 中所有 28 个案例的交互作用(LV* minority % VAP)考虑在内的时候,系数为正(b =0.52)但不显著。

这些虚假变量无法控制许多额外的选举制变化的因素(限制选票的程度、CV 或 LV 选举产生的所有席位的比例、辖区使用 CV 或 LV 的年份,等等)。篇幅所限,我们无法将所有这些因素都包括在内,因而也就无法确切地得出结论,美国实行的 LV 比 CV 产生更成比例的少数族群的代表。

最后,我们应当注意到,我们有理由料想到,有待竞选的席位数量会影响到任何选举制度的比例性(Lijphart 1994)。我们的确将这一因素作为自变量包含在最初的模型中,但效果并不显著。由于我们的目的是复现其他并未包含这一因素的研究模型,所以本文所报告的模型并不包含这一因素。在这些选举制中,参选的席位数量存在有限的变化,所以很难评价这个变量的影响。

讨 论

我们的数据提供证据表明,采用限制或累积投票制修正地方不分区选举制,有望产生与单一席位选区制下所发现的水平相当的少数族群的代表性。这一发现应该会鼓舞这样一些人,他们致力于促进少数族群的代表性,而又不依赖于基于种族划分选区的苛刻程序。先前研究业已证明,少数族群的确在修正后的选举方案下赢得席位。我们的目的是要验证 CV/LV 下的席位—人口关系如何与其他方案下所产生的关系进行比较。对非洲裔美国人而言,CV/LV 选举制下的代表性可以顺利地与单一席位选区制下所获得的代表性进行比较,而且比未修正的不分区选举制下的代表性更成比例。对于本文所研究的社区中拉美人而言,情况并非如此。

我们对关于拉美地区的研究结果提出如下问题:为什么 CV 选举结果更不成比例。在本文所研究的案例中,我们的估计所反映出的比例不足,可能与以下三方面因素的作用相关:(1)拉美裔候选人选拔不足;(2)对拉美裔达到法定投票年龄的人口的普查测量(这些模型的关键自变量),与实际参加投票的拉美裔人口之间存在实质差距;(3)占主导地位的少数族群为拉美人的许多地区所采用的 CV 方案本身具有较高的排斥临界值。

在我们的案例中,大量的德克萨斯州的 CV 选举($N=33$)实际上拥有拉美人口,却并未实现拉美人的代表性。在其中的 19 宗案例中,市政员指出,没有少数族群候选人参加竞选。同样,只有两宗案例中非洲裔美国人构成主导地位的少数派群体,但没有非洲裔候选人谋得公职。数据显示,提名问题对拉美辖区更重要,由于这个原因,提名不足可以部分解释为什么在拉美人占主导地位的辖区,席位和人口份额之间缺少一种线性关系。

另外还有 14 宗案例,拉美裔候选人在拉美裔人口明显占主导地位的地区被击败。第二个重要因素,影响到我们估计修正后的不分区选举制下的拉美人的代表性的能力,这一因素就是运用普查测量达到法定投票年龄的人口。在本分析所包含的

大多数拉美地区中，拉美裔参与投票者比白人参与者要少很多。[14] 少数族群法定选举人口的百分比数据，可能会产生出比过去所运用的登记投票人数或出席投票人数的数据所产生的估计斜率更低，因为后者更精确地反映了少数族群的选举力量。同样，法定投票人口的普查测量与实际出席率之间的差距，相对于非洲裔美国人而言，可能要比相对于拉美人而言更小。当拉美人出席投票的人数数据包含在一个局限于14 个德克萨斯州的社区（在这些社区中，这些数据是可资利用的，并且拉美裔候选人在 1995 年谋得了公职）时，拉美裔席位—出席人数的关系斜率为 1.22（R^2 = 0.30），并且这一结果具有统计显著性。[15] 这表明，当拉美投票者被动员起来时，拉美裔候选人在累积投票制下可以做得很好。

第三个因素是加重人口与出席投票人数之间的差距。德克萨斯州的大部分地区，每一次单一选举往往只设两到三个通过累积投票制竞选的席位。本文的分析案例，除两个以外其余的所有拉美裔案例均来自德克萨斯州。这意味着，德克萨斯州的排斥临界值通常为 25% 或 30% 。当排斥临界值如此高的时候，较低的少数族群出席投票人数便成为突出问题。

我们应当强调，解释低水平的拉美裔代表性的这些因素中，没有一种因素是采用修正后的不分区选举制度自动产生的结果。毋宁说，限制因素就在于：一些 CV 方案如何加以设计，团体如何利用这种选举制度。如果一次选举中根据累积投票制竞选的席位不超过两个，那么任何少数族群都将难以赢得席位，除非他们以团体形式投票，并控制相当规模的法定选举人口份额（接近或超过 33%），并且动员投票者参加投票，直到投票人数与多数派团体的投票者持平甚至超过。

然而，如果选举方案产生的排斥临界值，并未超过少数派团体的选举力量，并且少数派政治团体有良好的组织，以至于能够选拔候选人甚或对提名有所控制，并且或者动员投票者将所有选票投给某一（些）候选人，那么，与累积投票制和限制投票制相关的许多策略负担就能克服。我们的结果表明，这些负担显然能够被克服，而且 CV/LV 方案的确有助于少数派团体在易实现的条件下获得成比例的描述性代表，同时又避免使用具有种族意识的选区划分。我们的结果表明，与累积投票制相关的许多策略负担和合作要求，很容易被少数族群候选人克服。

---

⑭ 在其中的两个辖区内，少数族群参与投票的人数实际上超出英裔美国人的投票人数，而且少数族群候选人成功当选。投票人数的数据来自 Brischetto and Engstrom，1998。

⑮ 德克萨斯州使用累积投票制的地区，只能得到 1995 年的种族团体参加投票人数的数据。当年，在 15 次德克萨斯州累计投票制选举中，拉美裔候选人谋得公职，登记投票人数的数据报给其中的 14 个地区。因此，我们无法运用所有 96 宗案例来准备相似的估计。这些数据来自 Brischetto and Engstrom，1998。



Amy, Douglas. 1993. *Real Choices, New Voices.* New York: Columbia University Press.

Brace, Kimball, Bernard Grofman, Lisa Handley, and Richard Niemi. 1988. "Minority Voting Equality: The 65 Percent Rule in Theory and Practice." *Law and Policy* 10 (1): 43-62.

Brischetto, Robert. 1995. "The Rise of Cumulative Voting." *Texas Observer*, 28 July, 6.

Brischetto, Robert, and Richard Engstrom. 1998. "Cumulative Voting and Latino Representation: Exit Surveys in Fifteen Texas Communities." *Social Science Quarterly.*

Bullock, Charles S. 1994. "Section 2 of the Voting Rights Act, Districting Formats, and the Election of African-Americans." *Journal of Politics* 56: 1098-1105.

Cole, Richard, and Delbert Taebel. 1992. "Cumulative Voting in Local Elections: Lessons from the Alamagordo Experience." *Social Science Quarterly* 73: 194-201.

Cole, Richard, Delbert Taebel, and Richard Engstrom. 1990. "Cumulative Voting in a Municipal Election: A Note on Voter Reactions and Electoral Consequences." *Western Political Quarterly* 43(4): 191-199.

Cox, Gary. 1997. *Making Votes Count.* Cambridge: Cambridge University Press.

Engstrom, Richard. 1993. "The Single-Transferable Vote: An Alternative Remedy for Minority Vote Dilution." *University of San Francisco Law Review* 27 (4): 781-813.

Engstrom, Richard, and Michael McDonald. 1981. "The Election of Blacks to City Councils: Clarifying the Impact of Electoral Arrangements on the Seats-Population Relationship." *American Political Science Review* 75 (2): 344-354.

Guinier, Lani. 1991. "No Two Seats: The Elusive Quest for Political Equality." *Virginia Law Review* 77: 1414-1514.

Guinier, Lani. 1994. *Tyranny of the Majority: Fundamental Fairness in Representative Democracy.* New York: Free Press.

Johnson, Richard. 1979. *Political, Electoral, and Spatial Systems.* Oxford: Oxford University Press.

Keech, William, and Michael Sistrom. 1994. "North Carolina." In *Quiet Revolution in the South: The Impact of the Voting Rights Act*, ed. Chandler Davidson and Bernard Grofman. Princeton: Princeton University Press.

Lakeman, Enid. 1970. *How Democracies Vote.* London: Faber and Faber.

Lijphart, Arend. 1994. *Electoral Systems and Party Systems.* New York: Oxford University Press.

Massey, Douglas, and Nancy Denton. 1987. "Trends in Residential Segregation of Blacks, Hispanics, and Asians: 1970-1980." *American Sociological Review* 52 (6): 802-825.

Pildes, Richard, and Kristen Donoghue. 1995. "Cumulative Voting in the United States." *University of Chicago Legal Forum*, 241-312.

Polinard, Jerry, Robert Wrinkle, and Tomas Longoria. 1991. "The Impact of District Elections on the Mexican American Community: The Electoral Perspective." Social Science Quarterly 72 (3): 608-714.

Sass, Tim R., and Stephen Mehay. N. d. "Minority Representation, Election Method, and Policy Influence." Department of Economics, Florida State University. Unpublished Manuscript.

Still, Edward. 1984. "Alternatives to Single-Member Districts." In *Minority Vote Dilution*, ed. Chandler Davidson. Washington, DC: Howard University Press.

Still, Edward. 1991. "Voluntary Constituencies: Modified At-Large Voting as a Remedy for Minority Vote Dilution in Judicial Elections." *Yale Law and Policy Review* 9(2): 354-369.

Swain, Carol. 1993. *Black Faces, Block Interests: The Representation of African-Americans in Congress.* Cambridge: Harvard University Press.

Taebel, Delbert. 1978. "Minority Representation on City Councils: The Impact of Structure on Blacks and Hispanics." *Social Science Quarterly* 59 (1): 142-152.

Thomas, James, and William Stewart. 1998. *Alabama Government and Politics.* Lincoln: University of Nebraska Press.

Vedlitz, Arnold, and Charles Johnson. 1982. "Community Segregation, Electoral Structure, and Minority Representation." *Social Science Quarterly* 63 (4): 729-736.

Welch, Susan. 1990. "The Impact of At-Large Elections on the Representation of Blacks and Hispanics." *Journal of Politics* 52 (4): 1050-1076.